



# Détermination de la structure par RMN d'une protéine impliquée dans la biosynthèse de centres [Fe-S]: SufA

Nicolas Duraffourg

## ► To cite this version:

Nicolas Duraffourg. Détermination de la structure par RMN d'une protéine impliquée dans la biosynthèse de centres [Fe-S]: SufA. Biophysique [physics.bio-ph]. Université Joseph-Fourier - Grenoble I, 2006. Français. NNT: . tel-00200198

**HAL Id: tel-00200198**

**<https://theses.hal.science/tel-00200198>**

Submitted on 20 Dec 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE JOSEPH FOURIER – GRENOBLE I  
SCIENCES & GEOGRAPHIE  
ÉCOLE DOCTORALE CHIMIE ET SCIENCES DU VIVANT

# THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ JOSPEH FOURIER**  
GRENOBLE I

**Discipline** : Chimie Physique Moléculaire et Structurale

Présentée et soutenue publiquement  
par

**Nicolas DURAFFOURG**

le 21 décembre 2006

Détermination de la structure par RMN d'une  
protéine impliquée dans la biosynthèse de  
centres [Fe-S] : SufA

---

**JURY :**

Dr. Jean-Marc LATOUR	Président
Dr. Françoise GUERLESQUIN	Rapporteur
Dr. François PENIN	Rapporteur
Dr. Dominique MARION	Co-Directeur de thèse
Prof. Marc FONTECAVE	Directeur de thèse

N°

Thèse préparée aux laboratoires de Chimie et Biochimie des Centres Redox Biologiques, DRDC-CB (UMR 5047 CEA-CNRS-UJF), de RMN - IBS « Jean-Pierre Ebel » (UMR 5075 CEA-CNRS-UJF)

## *Remerciements*

*Je tiens à remercier tous les membres du jury d'avoir accepté de juger ce travail.*

*Je remercie le Docteur Françoise Guerlesquin ainsi que le Docteur François Penin d'avoir accepté d'être rapporteur de ce travail ainsi que pour leurs conseils avisés.*

*Je remercie le Docteur Jean Marc Latour d'avoir accepté d'examiner ce travail et de présider le jury de ma thèse.*

*Je remercie le Docteur Dominique Marion, co-directeur de cette thèse de m'avoir accueilli au sein de son groupe, pour sa confiance et ses encouragements au cours de ces années. Je reste toujours admiratif des explications métaphoriques que tu m'as quelques fois distillées me laissant croire que certains aspects de la RMN pouvaient être simples... (en général, ça ne durait pas longtemps).*

*Je remercie le Professeur Marc Fontecave et lui exprime ici toute mon estime de m'avoir accueilli au sein du laboratoire CBCRB ainsi que de la confiance qu'il m'a accordée dès mon arrivée. Cette aventure n'aurait jamais eu lieu sans ton soutien et ton écoute dans les moments délicats.*

*Le présent travail a été réalisé entre le laboratoire CBCRB du CEA Grenoble auquel je suis rattaché et le laboratoire LRMN de l'institut de Biologie structurale. Je voudrais tout d'abord remercier Jean-Pierre Simorre, Directeur du laboratoire LRMN, d'avoir accepté ma présence durant ces dernières années ainsi que tes conseils avisés.*

*Je voudrais maintenant remercier deux personnes pour qui j'ai une très grande estime. Tout d'abord Le Docteur Laurence Blanchard, sans laquelle je n'aurais jamais fait ce travail de thèse. Merci pour ta gentillesse durant toutes ces années, ta patience et ta grande pédagogie. Merci pour tes encouragements et les discussions scientifiques... ou non. Je te souhaite une belle réussite professionnelle dans tes nouvelles fonctions. Tu es quelqu'un de bien ! soit « cool... »*

*La deuxième personne que je ne veux pas oublier est le Docteur Klaartje Houben. Tu es celle qui m'a le plus aidée durant ces années et je t'en suis extrêmement reconnaissant. Merci de m'avoir initié au calcul de structure avec notre ami « ARIA » ainsi que pour ta grande patience et ton aide permanente. J'admire ta grande connaissance scientifique et reste surtout admiratif de tes capacités linguistiques, moi qui ne suis pas vraiment doué dans ce domaine. Et surtout, merci pour ton amitié quotidienne ainsi que tous les bons*

*moments partagés en montagne. Espérons qu'il y en aura d'autres. Je voudrais te souhaiter une grande réussite scientifique et personnelle.*

*Un grand merci à Paul Schanda, brillant Doctorant Autrichien, qui comme Klaartje est un expert linguistique mais surtout un grand sportif avec qui j'ai partagé d'excellents moments que ce soit au laboratoire ou en montagne. Merci pour ta gentillesse et ton aide durant cette thèse. Vivement qu'on retourne se faire une « petite » balade...*

*Merci à toi, Adrien, de m'avoir supporté dans ton bureau et d'avoir supporté les sollicitations fréquentes à propos de petits problèmes informatiques... ainsi que pour tes conseils avisés à propos des calculs de structures. Merci pour les pauses café.*

*Je remercie Catherine, Béate, Isabelle, Bernhard, Martin, Jérôme, Pierre, du laboratoire LRMN pour leur sympathie et disponibilité.*

*Merci aussi à tous les étudiant(e)s et post-doc de ce laboratoire pour leur gentillesse et leur aide : Hélène, Rodolpho, Ewen, Thomas, Guillaume, Rémy, Julien ainsi que Denis.*

*Du côté du laboratoire CBCRB je voudrais particulièrement remercier le Docteur Sandrine Ollagnier de Choudens de m'avoir permis de travailler sur cette protéine et de m'avoir toujours fourni efficacement les échantillons nécessaires. Je voudrais aussi te remercier pour ta grande rigueur au cours de la lecture de ce document et pour tes précieux conseils. Tu es donc la troisième fée qui s'est penchée sur moi et sans laquelle ce travail ne serait pas ce qu'il est.*

*Je voudrais aussi remercier le Docteur Vincent Nivière avec qui j'ai débuté ce travail de thèse. Merci pour ta pédagogie et de m'avoir initié à la purification de protéine.*

*Je voudrais, ici, associé le Docteur Jacques Gaillard, pour qui j'ai le plus profond respect. Merci à toi Jacques d'avoir accepté de me diriger au début de cette thèse. Merci pour ton humanité et tout ce que tu m'as appris. Je te souhaite de bien profiter de ta nouvelle vie.*

*Je remercie le Docteur Stéphane Ménage, avec qui j'ai commencé cette aventure dans le laboratoire de chimie et avec lequel je partage un certain goût de la répartie... j'avoue que tu es un peu meilleur que moi mais je progresse. Bref, j'espère que l'on retravaillera ensemble.*

*Merci à tous les autres membres du laboratoires pour toutes les discussions autour du café comme pour les franches rigolades et les bonnes eng... .. Bref, merci pour cette bonne tranche de vie au laboratoire. Je ne regrette pas d'être là et compte bien y rester encore un peu.*

*Merci à tous les Thésards et post-Doct du laboratoire passé et présent pour leur bonne humeur, leur sympathie et les bonnes discussions pas toujours très sérieuses. Vu le nombre, si je dois en citer une, c'est toi Maïté de m'avoir fourni des échantillons et pour les discussions intéressantes que nous avons pu avoir sur le sujet.*

*Merci à toi Olivier, pour ton amitié et ton soutien permanent. Pour ta présence dans les moments forts, voir difficiles de ces dernières années, aussi bien professionnels que personnels.*

*Pour terminer, je voudrais remercier le CEA pour son accompagnement permanent au cours de cette longue aventure de formation, pour la confiance qui m'a été témoigné aussi bien sur les centres CEA de Valduc et de Grenoble. Pour les mêmes raisons je remercie l'université de Bourgogne et l'université Joseph Fourier de Grenoble.*

*Je ne pourrais pas être digne si j'oubliais de remercier ma famille, parents et frères, qui m'ont toujours encouragé et soutenu au cours de toutes ces années. Bien entendu, si il y une personne que je ne peux oublier c'est toi, Alexandre, mon garçon qui ne m'a pas beaucoup vu ces derniers temps. Je te dédie cette thèse et suis fier de toi. Merci à toi, Emmanuelle pour ton soutien, ton aide et tes encouragements au cours de cette rédaction, d'avoir supporté quelques nuits blanches et bien entendu pour tout ce que nous avons partagé...*

## Avant-propos

Cette thèse est l'aboutissement d'une démarche entamée voilà maintenant presque dix ans. Suite à l'obtention en 1999 d'une Maîtrise en Sciences et Techniques " Contrôle et Analyse Chimique " de l'Université de Bourgogne, et après avoir obtenu le DEA " Chimie-Physique Moléculaire et Structurale " l'année suivante, j'ai intégré le laboratoire Chimie Biologie des Centres Rédox Biologiques du Professeur Marc Fontecave au CEA Grenoble, au sein duquel j'assume la maintenance et développement du spectromètre de résonance magnétique nucléaire (RMN) Brüker Avance 300, ainsi que diverses responsabilités administratives.

En accord avec ma formation, je me suis naturellement orienté vers l'activité chimie du laboratoire, principale utilisatrice du spectromètre RMN, et plus particulièrement vers l'étude par RMN du proton ( $^1\text{H}$ ) et du carbone ( $^{13}\text{C}$ ) de diastéréoisomères de complexes de Fer. Une des spécificités du laboratoire est d'associer chimistes et biologistes afin de mieux appréhender certains mécanismes biochimiques fondamentaux mais aussi de développer de nouveaux catalyseurs chimiques capables d'effectuer, par exemple de l'oxydation énantiosélective ; cette collaboration passe pour les chimistes par la synthèse de composés inorganiques via une approche biomimétique originale. Dans ce cadre Ménage S et collaborateurs ont synthétisé un certain nombre de complexes mononucléaires et binucléaires à Fer. Associé à cette problématique, j'ai débuté mon étude par la synthèse de complexes mononucléaires de Fer par auto-assemblage avec des ligands organiques polyaromatiques du type bipyridine, phénantroline... L'objectif était de pouvoir discriminer les deux énantiomères  $\Delta$  et  $\Lambda$ , issus de la chiralité au métal, via la formation de diastéréoisomères, que l'on espérait pouvoir séparer. Cette étude reposait essentiellement sur l'analyse de spectres 1D, parfois 2D (2D COSY), comme la plupart des analyses RMN effectuées par les chimistes. A l'issue de mon DEA, j'ai émis le vœu de poursuivre en thèse, mais il me semblait alors plus pertinent, étant rattaché à la Direction des Sciences du Vivant du CEA, de m'orienter vers l'analyse d'entités biologiques, telles que les protéines et plus précisément les métallo-enzymes étudiées au laboratoire. Ceci devrait me permettre d'élargir mes compétences, et par conséquent en apporter de nouvelles au laboratoire.

## Abréviations

1D, 2D, 3D	Spectre à 1, 2, 3 dimensions
ADN	Acide désoxyribo nucléique
ARIA	Ambiguous Restraints in Iterative Assignment
aniso	anisotropique
ARN	Acide ribonucléique
CaCl <sub>2</sub>	Chlorure de calcium
CSA	Anisotropie du déplacement chimique
CSI	Index des déplacements chimiques
CNS	Cystallography and NMR Systeme
<sup>n</sup> D <sub>IS</sub>	Couplage dipolaire à n liaisons entre le spin I et le spin S
Da	Dalton : poids moléculaire (1 Da ≈ 1g/mol)
EDTA	Ethylène diamine tétra-acétate
ENH	Electrode normale à hydrogène
H-bound	Liaison hydrogène
HCl	Acide chlorhydrique
HSQC	Heteronuclear Single Quantum Coherence
Hz	Hertz
IPTG	IsoPropylThioGalactoside
IsC	Iron sulfur cluster
iso	Isotrope
<sup>n</sup> J <sub>IS</sub>	Couplage scalaire à n liaisons entre le spin I et le spin S
K <sub>d</sub>	Constante de dissociation
mA	milliampère
MgCl	Chlorure de magnésium
mg	milligramme
min	minute
mL	millilitre
mV	millvolt
μM	micromolaire
NaCl	Chlorure de sodium

NH <sub>4</sub> Cl	Chlorure d'ammonium
NiNTA	Nickel NitriloTriAcetic acid
nm	nanomètre
nOe	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Enhancement Spectroscopy
PAGE	Polyacrylamide gel electrophoresis
PCR	polymerase Chain Reaction
PDB	Protein Data Bank
ppm	partie par million
RDC	Residual Dipolar Coupling
RMN	Résonance Magnétique Nucléaire
RPE	Résonance Para-Electronique
rpm	tour par minute
SDS	Sodium DodécylSulfate
SOR	SuperOxydeRéductase
Suf	protéine Suf (pour Sulfure)
TALOS	Torsion Angle Likelihood Obtained from Shift and sequence similarity
TROSY	Transverse Relaxation-Optimized Spectroscopy
UV	Ultra-Violet
V	Volt



# TABLE DES MATIERES

---

INTRODUCTION.....	1
-------------------	---

PREMIERE PARTIE - LA PROTEINE SUFA .....	
--	--

---

## **CHAPITRE I – BIOSYNTHESE DES CENTRES [Fe-S] 3**

INTRODUCTION .....	3
--------------------	---

I-LES DIFFÉRENTS CENTRES [Fe-S] .....	4
---------------------------------------	---

I.1-Structure atomique des centres Fe-S.....	4
--	---

I.1.1-Les centres [1Fe] .....	5
-------------------------------	---

I.1.2-Les centres [2Fe-2S].....	6
---------------------------------	---

I.1.3-Les centres [4Fe-4S].....	8
---------------------------------	---

I.1.4-Les centres [3Fe-4S].....	9
---------------------------------	---

I.1.5-Les centres complexes.....	10
----------------------------------	----

II-FONCTIONS DES PROTÉINES [Fe-S] .....	11
---	----

II.1-Les centres [Fe-S] dans le transfert d'électrons .....	11
---	----

II.2-L'activité enzymatique des centres [Fe-S] .....	12
--	----

III-BIOSYNTHESE DES CENTRES [Fe-S] .....	14
--	----

III.1-Introduction .....	14
--------------------------	----

III.2-Le système SUF .....	16
----------------------------	----

III.2.1-Analyse physiologique.....	16
------------------------------------	----

III.3-Les protéines Suf.....	17
------------------------------	----

III.3.1-Le complexe SufSE .....	17
III.3.2-Le complexe SufBCD .....	20
II.3I.3-La protéine SufA .....	20
III.4-Mécanisme d'assemblage de centre et transfert.....	23
III.4.1-L'assemblage de centre .....	23
III.4.2-Le transfert de centre .....	25

## **CHAPITRE II - ETUDE PAR RMN D'UNE PROTEINE DE GRANDE TAILLE ..... 27**

### **I-ÉTUDE PAR RMN D'UNE PROTÉINE DE MASSE MOLÉCULAIRE ÉLEVÉE..... 29**

I.1-Limites imposées par la méthode de la RMN .....	29
I.1.1-Limites liées aux propriétés de l'échantillon biologique .....	29
I.1.2-Limites liées à la taille de la protéine étudiée.....	29
I.2-Les principales avancées visant à repousser les limites de la RMN.....	30
I.2.1-Les spectromètres.....	30
I.2.2-Les cryosondes.....	31
I.2.3-Le marquage isotopique des échantillons pour une RMN multidimensionnelle.....	31
I.2.4-La deutération .....	32
I.2.5-La spectroscopie TROSY .....	35

### **II-ETAPES DE L'ÉTUDE PAR RMN DE PROTÉINE DE GRANDE TAILLE ..... 37**

II.1-Expression de protéines recombinantes et marquées pour la RMN.....	37
II.1.1-Optimisation d'un système d'expression .....	38
II.1.2-Le milieu de production .....	39
II.1.3-Les techniques de marquage isotopique.....	39
II.1.4-Préparation de l'échantillon RMN.....	40
II.2-Attribution des raies de résonances d'une protéine marquée.....	41
II.2.1-Attribution de la chaîne principale et des C $\beta$ .....	42

II.2.2-Attribution des chaînes latérales .....	46
II.3-Recueil des contraintes structurales.....	48
II.3.1-Les contraintes mesurables classiques .....	48
II.3.2-Les contraintes orientationnelles .....	51
II.3.2.1- Interaction dipolaire .....	52
II.3.2.1.1- Aspect théorique .....	52
II.3.2.1.2-Les bactériophages filamenteux.....	55
III-MODÉLISATION MOLÉCULAIRE .....	56
III.1-Introduction.....	56
III.2-Le calcul de structure.....	57
III.2.1-Le champ de force .....	58
III.2.2-Méthode de calcul – ARIA1.2 .....	60
IV-ETUDE DE DYNAMIQUE .....	62
<b>CHAPITRE III - PRODUCTION ET ATTRIBUTION DE LA</b>	
<b>PROTEINE SufA .....</b>	<b>67</b>
I-PRÉPARATION DE L'ÉCHANTILLON DE LA PROTÉINE SufA .....	67
II.1-Résultats de production de la protéine SufA.....	68
II-CARACTÉRISATION PRÉLIMINAIRE DE SufA .....	68
III.1-Caractéristiques de SufA d' <i>E. coli</i> .....	69
III.2-Prédiction de structures secondaires de SufA.....	70
III.3-Echantillon RMN .....	71
III-ATTRIBUTION DES RÉSONANCES .....	72
III.1-Stratégie d'attribution des résonances de la chaîne principale .....	73
III.2-Attribution séquentielle et CSI.....	74
III.3-Analyse de mutants .....	76

III.4-Analyse de la protéine SufA avec son centre [Fe-S].....	78
III.5-Attribution des chaînes latérales .....	82
III.6-Attribution des aromatiques.....	84
III.7-Note d'attribution .....	86
III.8- Échange de protons amides.....	86

## CHAPITRE IV - DETERMINATION DE LA STRUCTURE DE SUFA..... 88

I-CONSTRAINTES DE DISTANCES .....	88
I.1-Extraction des nOe courtes portées.....	90
I.2-Extraction des nOes inter-brins $\beta$ et moyenne portée .....	90
I.3-Extraction des nOe longue portée.....	92

II-CONSTRAINTES D'ANGLES DIÈDRES.....	96
---------------------------------------	----

III-CONSTRAINTES ORIENTATIONNELLES .....	96
III.1-Préparation du milieu cristal liquide et des échantillons .....	97
III.1.1-SufA dans les bactériophages <i>Pfl</i> .....	97
III.1.2-Mesure de couplages dipolaires résiduels le long de la chaîne polypeptidique.....	98
III.1.2.1-Principe de la mesure.....	98
III.1.2.2-Expériences enregistrées sur SufA .....	99
III.1.2.3-Extraction des couplages dipolaires résiduels.....	99

IV-CONSTRAINTES DE LIAISONS HYDROGÈNE .....	100
---	-----

IV- CALCUL DE STRUCTURE DE SufA .....	101
IV.1- Calcul avec ARIA/CNS.....	101

V- ANALYSE DE LA DYNAMIQUE DE SufA .....	108
--	-----

## CHAPITRE V- DISCUSSION.....110

## SECONDE PARTIE - LA SOR DE *TREPONEMA PALLIDUM*.....

..... 120

I-CONTEXTE BIOLOGIQUE ..... 120

II-MATÉRIELS ET MÉTHODES..... 123

III-RÉSULTATS ET DISCUSSION ..... 124

IV-PRODUCTION DE SOR MARQUÉE DE *T. pallidum*..... 126

IV.1-Production de la protéine SOR marquée ..... 126

V-RÉSULTATS DE PRODUCTION DE LA PROTÉINE SOR ..... 127

V.1-Caractérisation de la SOR de *Treponema pallidum* ..... 128

V1.1-Caractéristique biochimique..... 128

V.1.2-Caractéristiques spectroscopiques ..... 129

V.1.2.1-Spectroscopie UV-visible..... 129

V.1.2.2-Spectroscopie de masse..... 129

VI-ANALYSE RMN ..... 130

## TROISIEME PARTIE - MATERIELS ET METHODES.....

---

### **CHAPITRE I : PRODUCTION DE LA PROTEINE SOR DE *T.***

#### ***PALLIDUM* ..... 136**

##### I-MATÉRIEL BIOLOGIQUE ..... 136

I.1-souches et plasmides ..... 136

I.2-Milieu de culture ..... 136

I.3-Culture de la souche bactérienne et surexpression de la SOR..... 136

##### II-TECHNIQUES DE BIOCHIMIE ..... 137

II.1-Préparation d'extraits cellulaires ..... 137

II.1-Traitement des extraits solubles ..... 137

II.3-Purification des protéines..... 138

II.3.1-Chromatographie par tamisage moléculaire sur gel ACA 54 ..... 138

II.3.2-Chromatographie par résine échangeuse d'anions Uno Q ..... 138

##### III-ANALYSES BIOCHIMIQUES ..... 138

III.1-Electrophorèse de protéines : SDS-PAGE..... 138

III.1.1-Préparation des échantillons..... 139

III.1.2-Conditions de migration et révélation des protéines ..... 139

III.2-Dosage standard ..... 140

### **CHAPITRE II : SUREXPRESSION ET PURIFICATION DE LA**

#### **PROTEINE SUFA ..... 141**

##### I-MATÉRIEL BIOLOGIQUE ..... 141

I.1-Clonage du gène SufA et construction du plasmide d'expression ..... 141

II-PRODUCTION ET PURIFICATION DE LA PROTÉINE SufA SIMPLEMENT ET DOUBLEMENT MARQUÉE PAR LES ISOTOPES $^{15}\text{N}/^{13}\text{C}$ .....	142
II.1-Production .....	142
II.2-Purification .....	143
II.2.1-Extraction .....	143
II.2.2-Purification .....	144
II.3-Reconstitution.....	144

<b>CHAPITRE III : SPECTROSCOPIE RMN ET CALCUL DE STRUCTURE .....</b>	<b>145</b>
I-SPECTROSCOPIE RMN .....	146
II-CALCUL DE STRUCTURE .....	154
<b>REFERENCES BIBLIOGRAPHIQUES .....</b>	<b>156</b>

---

Comme nous le verrons, les protéines à centre fer-soufre ([Fe-S]) contrôlent un grand nombre de fonctions biologiques. Récemment, un déficit de centre [Fe-S] a été reconnu comme une des caractéristiques biochimiques majeures de pathologies neurodégénératives comme l'ataxie de Friedreich, liée à un défaut d'expression d'une protéine mitochondriale, la frataxine, qui pourrait jouer le rôle de donneur de fer dans la biosynthèse des centres [Fe-S] [1]. Le dysfonctionnement de cette protéine entraînerait un défaut de fabrication de centres [Fe-S] avec pour conséquence l'accumulation de fer dans la mitochondrie. Cette accumulation de fer a pour effet de favoriser le stress oxydant au sein de la cellule. Ce phénomène biochimique, affectant les organismes vivant en aérobiose, trouve son origine dans les imperfections de la chaîne respiratoire au sein de la cellule pour aboutir à la formation d'entités dites espèces activées de l'oxygène (EAO), dotées de propriétés oxydantes parmi lesquelles on trouve l'anion superoxyde  $O_2^{\bullet -}$ , le peroxyde d'hydrogène  $H_2O_2$ , et le radical hydroxyle  $OH^{\bullet}$  [2].

Ce phénomène de stress oxydant doit être considéré comme un processus global où les métabolismes du dioxygène et du fer sont étroitement liés. D'une part, la présence de fer libre dans le cytoplasme a pour effet de favoriser la génération du radical  $OH^{\bullet}$ , via la réaction de Fenton, et d'autre part les EAO formées sont susceptibles de dégrader les centres [Fe-S], libérant ainsi du fer libre dans la cellule. Dans ce contexte, le laboratoire étudie deux processus biochimiques, le premier étant lié à la détoxification de l'anion superoxyde  $O_2^{\bullet -}$  par la protéine superoxyde réductase (SOR), le second étant lié à la biosynthèse des centres [Fe-S] contrôlé par différents systèmes protéiques chez *E. coli*, organisés en opérons. Parmi ces systèmes, on trouve le système SUF (Sulfur mobilisation). Cet opéron, composé de six protéines, est formé de deux complexes protéiques qui seront décrits ultérieurement et d'une protéine particulière, SufA, dont le rôle est de pré-assembler un centre [Fe-S] avant de le transférer à une protéine cible acceptrice. Si on commence à mieux cerner le mécanisme par lequel ces différents systèmes fonctionnent, un certain nombre de questions restent en suspens à ce jour. Il en va ainsi sur la façon dont la protéine SufA préassemble le centre [Fe-S] et le transfère. Pour essayer de le comprendre, il nous a semblé important de déterminer la structure de cette protéine dont aucune n'était connue au début de ce travail de thèse, mis à

---

<sup>1</sup> Puccio, H. & Koenig, M. Friedreich ataxia: a paradigm for mitochondrial diseases. *Curr Opin Genet Dev* **12**, 272-7 (2002).

<sup>2</sup> Imlay, J. A. (2006) Iron-sulphur clusters and the problem with oxygen, *Molecular microbiology* **59**, 1073-1082.



part la structure cristallographique de la protéine homologue, IscA d'*E. coli* [<sup>3</sup>]. Il me semble important de mentionner ici qu'au cours de ce travail de thèse la structure cristallographique de la protéine SufA d'*E. coli* a été publiée [<sup>4</sup>]; cependant, les différences observées, entre cette structure et les résultats préliminaires que nous avons, nous ont encouragé à poursuivre. Cette étude fait l'objet de la première et principale partie de cette thèse alors que la deuxième partie est dédiée à l'étude structurale de la protéine SOR de *Treponema pallidum*.

La biologie structurale constitue un domaine de recherche performant pour tenter d'élucider les mécanismes fonctionnels des biomolécules qui régissent les organismes. Ainsi les caractérisations structurales et dynamiques à l'échelle atomique des protéines et des acides nucléiques, informent sur leurs fonctions spécifiques. La **Résonance Magnétique Nucléaire** (RMN) est une des principales techniques expérimentales permettant d'apporter de telles informations. Dans le but d'améliorer la compréhension du mode de fonctionnement des protéines SufA et SOR, de structures inconnues au début de nos recherches, nous avons donc entrepris leurs études structurales et dynamiques par RMN. Une brève description de l'approche adoptée, dans cette discipline, pour l'étude de protéine de tailles élevées telle que SufA (2x14 KDa) est exposée dans la première partie.

---

<sup>3</sup> Cupp-Vickery, J. R., Silberg, J. J., Ta, D. T. & Vickery, L. E. Crystal structure of IscA, an iron-sulfur cluster assembly protein from *Escherichia coli*. *J Mol Biol* **338**, 127-37 (2004).

<sup>4</sup> Wada, K., Hasegawa, Y., Gong, Z., Minami, Y., Fukuyama, K., and Takahashi, Y. (2005) Crystal structure of *Escherichia coli* SufA involved in biosynthesis of iron-sulfur clusters: implications for a functional dimer, *FEBS letters* **579**, 6543-6548.

## Chapitre I

# Biosynthèse des centres [Fe-S]

## INTRODUCTION

Le fer est un élément indispensable à la vie, et sa carence dans un organisme vivant peut occasionner des maladies graves, parfois même létales : l'anémie chez les mammifères, la chlorose ferrique chez les végétaux [1].

Les protéines fer-soufre sont sans doute parmi les premiers catalyseurs que la nature ait utilisés. Il semblerait, en effet, que les conditions extrêmes régnant dans l'atmosphère à cette époque, aient favorisé une chimie du fer et du soufre, d'ailleurs très présents dans la soupe primitive [2]. Paradoxalement, ces entités ancestrales, et comme nous le verrons fondamentales dans le règne du vivant, n'ont été découvertes que très récemment. C'est en effet seulement à partir des années 1960 que plusieurs résultats expérimentaux tendaient à prouver l'existence de métalloprotéines à fer d'un type tout à fait nouveau [3]. En 1962, Mortenson et ses collaborateurs isolaient une protéine à fer non héminique de *Clostridium pasteurianum*, qui s'avère essentielle à la réaction d'oxydation du puryvate [4]. Les auteurs décidèrent de lui donner le nom de ferrédoxine (Fd), contenant du fer et intervenant dans les réactions d'oxydo-réduction.

Simultanément, une protéine est isolée à partir des chloroplastes d'épinard et, fait troublant, peut être remplacée par la ferrédoxine de *Clostridium pasteurianum* sans affecter sa fonction initiale [5]. De plus, avec la mise en évidence de soufre labile dans ces protéines [6], celles-ci deviennent les premiers représentants d'une nouvelle classe de protéines: les protéines fer-soufre ([Fe-S]). Elles se distinguent par la présence d'un site actif constitué par

un ou plusieurs atomes de fer liés entre eux par du soufre inorganique, l'ensemble étant le plus souvent lié à la protéine par les atomes de soufre de cystéines.

Les protéines fer-soufre ont aujourd'hui perdu une grande partie de leur mystère et il n'est sans doute pas de classe de métalloprotéines, à l'exception peut-être des protéines héminiques, qui ait fait l'objet d'investigations aussi poussées. Dans ce domaine, l'utilisation conjointe de techniques et d'approches provenant aussi bien de la biologie, de la chimie, que de la spectroscopie physique s'est avérée essentielle. En effet cette démarche a permis d'établir, partiellement du moins, des liens entre la nature du centre [Fe-S] considéré et le rôle physiologique qu'il assure au sein de la protéine. Par contre, la biosynthèse, c'est à dire la manière dont ils sont assemblés *in vivo*, n'est pas à ce jour complètement élucidée.

Ainsi après une brève présentation des caractéristiques générales des protéines [Fe-S], nous ferons un état des lieux concernant cette biosynthèse en nous appuyant plus particulièrement sur les systèmes étudiés au laboratoire.

## **I-LES DIFFÉRENTS CENTRES [Fe-S]**

Les nombreuses études réalisées sur la famille des protéines fer-soufre ont montré leur importance dans des processus essentiels à la vie, aussi divers que la photosynthèse, la fixation de l'azote, le métabolisme de l'hydrogène ou la respiration cellulaire. Une si grande variété fonctionnelle ne s'accompagne pas, comme nous pourrons le voir ci après, d'une aussi large diversité structurale. Qui plus est, il n'est pas évident d'établir une étroite relation entre ces deux aspects.

### **I.1-Structure atomique des centres Fe-S**

Les centres Fe-S, présents dans les sites actifs de protéines, sont des agrégats polymétalliques composés d'atomes de fer et de soufre agencés en stoechiométrie variable. Ils sont reliés le plus souvent à la chaîne polypeptidique par les atomes de soufre des cystéines,

mais aussi parfois par des atomes d'azote d'histidines ou des atomes d'oxygène de sérines ou d'aspartates [7]. Dans ces édifices, on distingue deux types d'atomes de soufre :

- les atomes labiles, notés S\* et appelés "soufre inorganique", qui sont liés uniquement aux atomes de fer.
- les atomes notés simplement S, qualifiés de "thiolates", qui sont liés à l'un des atomes de fer de l'agrégat et qui appartiennent à une cystéine de la chaîne polypeptidique.

On distingue plusieurs types de centre : [1Fe], [2Fe-2S], [3Fe-4S], [4Fe-4S], considérés comme des centres simples, par opposition aux centres complexes qui correspondent à des formes plus exotiques (centres Fe-S associés à d'autres groupes prosthétiques ou à d'autres métaux tels que le nickel, le molybdène ou encore le cuivre).

### I.1.1-Les centres [1Fe]

Il est constitué d'un atome de fer central dans une géométrie pseudo-tétraédrique, entouré de quatre atomes de soufre de résidus cystéine de la chaîne protéique (Figure 1) ; ce centre ne correspond pas à proprement parler à la définition d'un centre fer-soufre puisqu'il ne possède pas de soufre labile. On le rattache toutefois à la famille des protéines fer-soufre en raison de sa symétrie et de son environnement exclusivement soufré.

On rencontre ce type de centre dans les protéines de faible masse moléculaire (environ 6 KDa), les rubrédoxines issues principalement de bactéries anaérobies telle que *Desulfovibrio vulgaris* [8]. Oxydé, l'atome de fer est ferrique et dans un état haut spin ( $S=5/2$ ) alors que la réduction en fer ferreux provoque une légère distorsion du centre avec notamment une augmentation des distances Fe-S ; le fer est aussi dans un état haut spin ( $S=2$ ).

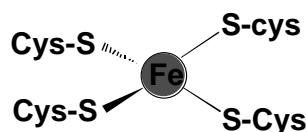


Figure 1 : Représentation schématique d'un centre [1Fe-1S]

centre [1Fe-1S]	type rubrédoxine
Etat rédox	Fe <sup>3+</sup> / Fe <sup>2+</sup>
Potentiel redox <sup>(a)</sup>	+20 à -100 mV

<sup>(a)</sup> Potentiels rédox donnés par rapport à l'ENH, à pH 7 à 25°C.

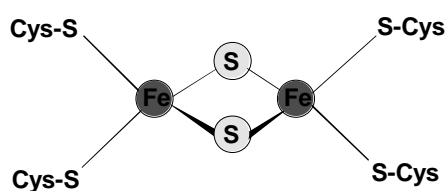
### I.1.2-Les centres [2Fe-2S]

De nombreuses études structurales de ce type de centre, représentés dans certaines ferrédoxines, ont permis de caractériser les séquences d'acides aminés impliqués dans les sites de liaison. Ainsi, certaines séquences consensus se sont révélées caractéristiques de l'insertion d'un type unique de centre fer-soufre. On distingue en effet quatre catégories de centres [2Fe-2S] :

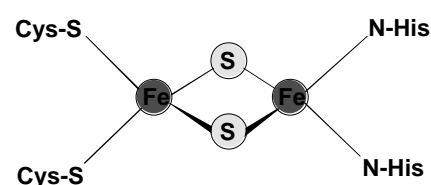
- Ferrédoxines de type plante avec une séquence consensus CX<sub>4</sub>CX<sub>2</sub>CX<sub>N</sub>C [<sup>9</sup>]. Cette séquence correspond au motif conservé contenant les cystéines supposées ou démontrées comme coordonnant le centre [Fe-S]. Elles sont isolées à partir d'organismes photosynthétiques parmi les végétaux supérieurs, les algues, les bactéries.
- Ferrédoxines de type vertébré (hydroxylase, adrénodoxine) avec une séquence consensus CX<sub>5</sub>CX<sub>2</sub>CX<sub>N</sub>C [<sup>10</sup>]. Elles sont présentes dans les bactéries non photosynthétiques et chez les mammifères.
- Ferrédoxines de type *Clostridium pasteurianum* avec une séquence consensus CX<sub>12</sub>CX<sub>31</sub>CX<sub>3</sub>C [<sup>11</sup>].

- Ferrédoxines de type Rieske avec une séquence consensus de type CXHX<sub>N</sub>CX<sub>2</sub>H. Le site actif est lié à la chaîne polypeptidique non pas par quatre cystéines mais par deux cystéines et deux histidines [12].

Le centre [2Fe-2S], de géométrie plane, est constitué de deux atomes de fer pontés par deux atomes de soufre labiles, et s'oriente perpendiculairement au plan formé par les atomes de soufre des cystéines ou des azotes des histidines pour les protéines de type Rieske (Figure 2). Quelle que soit la nature des ligands, le centre [2Fe-2S] existe dans deux états rédox, oxydé et réduit. Dans l'état oxydé, les atomes de fer sont tous les deux sous formes ferriques, de sorte que la charge portée par le site actif est +2 et le spin total, résultant du couplage antiferromagnétique entre les spins électroniques  $S=5/2$  de chacun des fers, est nul ( $S=0$ ). En revanche la réduction à un électron conduit à un centre constitué d'un fer ferrique et d'un fer ferreux, qui possède alors une charge résultante égale à +1 et spin total  $S$  égal à  $1/2$ .



A = type Ferrédoxine



B = type Rieske

Figure 2 : Représentation schématique d'un centre [2Fe-2S]

centre [2Fe-2S]	type Ferrédoxine	type Rieske
Etat d'oxydation	$\text{Fe}^{3+} / \text{Fe}^{3+}$ , $S=0$ , diamagnétique	$\text{Fe}^{3+} / \text{Fe}^{3+}$ , $S=0$ , diamagnétique
et Spin total	$\text{Fe}^{3+} / \text{Fe}^{2+}$ , $S=1/2$ , paramagnétique	$\text{Fe}^{3+} / \text{Fe}^{2+}$ , $S=1/2$ , paramagnétique
Potentiel redox <sup>(a)</sup> (+2/+1)	+70, -470 mV	-100, +400 mV

<sup>(a)</sup> Potentiels rédox donnés par rapport à l'ENH, à pH 7 à 25°C.

### I.1.3-Les centres [4Fe-4S]

Ces centres adoptent une géométrie de cube déformé où chacun des quatre atomes de fer est coordonné à trois atomes de soufre inorganique et un groupement thiolate d'une cystéine ligand (Figure 3). Il existe en fait une déformation du cube résultant de la tétravalence des atomes de fer qui occupent les sommets de l'édifice alternativement avec les atomes de soufres inorganiques.

Contrairement aux autres types de ferrédoxines, les centres [4Fe-4S] ne sont pas liés à une séquence consensus type. Au contraire, une étude bio-informatique [13] récente a montré qu'il existait, jusqu'à présent, une quarantaine de séquences différentes impliquées dans la chélation de ces centres.

Cette structure existe sous trois états d'oxydation différents, constituant ainsi deux couples rédox : les couples  $[4\text{Fe-4S}]^{+2/+1}$  et  $[4\text{Fe-4S}]^{+3/+2}$ . Les protéines contenant ce dernier couple sont appelées HiPIPs (High Potential Iron Protein) en raison de la valeur élevée de leur potentiel d'oxydoréduction. Il convient de noter que la plupart des protéines [4Fe-4S] sont isolées dans l'état  $[4\text{Fe-4S}]^{2+}$ , de spin total  $S=0$ , et par conséquent non détectable en RPE. On peut toutefois accéder à un centre réduit  $[4\text{Fe-4S}]^{+1}$  pouvant se décomposer en deux paires d'atomes de fer: une paire ferreuse ( $\text{Fe}^{2+}, \text{Fe}^{2+}$ ) et une paire dite de valence mixte ( $\text{Fe}^{2.5+}, \text{Fe}^{2.5+}$ ). Cette réduction conduit à un signal RPE situé à une valeur du tenseur  $g_{\text{moyen}}$  inférieur à 2 correspond à états de spin = 1/2. Les protéines HiPIPs, quant à elles, sont facilement repérables, puisqu'elles se manifestent dans l'état oxydé  $[4\text{Fe-4S}]^{+3}$  par un signal RPE dont la valeur du tenseur  $g_{\text{moyen}}$  est supérieur à 2 caractéristique d'un état de spin  $S=1/2$ .

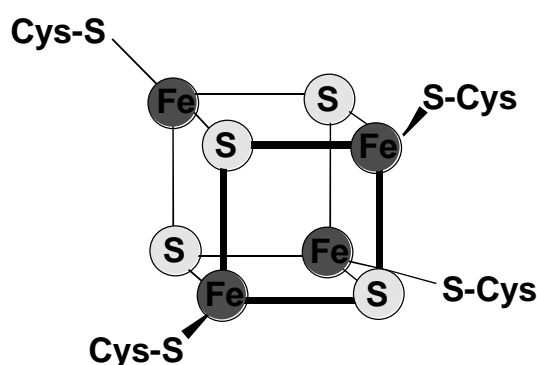


Figure 3 : Représentation schématique d'un centre [4Fe-4S]

<b>[4Fe-4S]<sup>+3</sup></b>	<b>[4Fe-4S]<sup>+2</sup></b>	<b>[4Fe-4S]<sup>+1</sup></b>
Fe <sup>2+</sup> , 3Fe <sup>3+</sup> S=1/2; g > 2	2Fe <sup>2+</sup> , 2Fe <sup>3+</sup> S=0	3Fe <sup>2+</sup> , Fe <sup>3+</sup> S=1/2, g < 2
Potentiel redox <sup>(a)</sup> (+3/+2) = +50 à +400 mV	Potentiel redox <sup>(a)</sup> (+2/+1) = -100 à -700 mV	

<sup>(a)</sup> Potentiels rédox donnés par rapport à l'ENH, à pH 7 à 25°C.

### I.1.4-Les centres [3Fe-4S]

Les ferrédoxines à centre [3Fe-4S], présentes dans le monde bactérien, possèdent un centre pseudo-cubique qui n'est pas relié à la protéine par des résidus organisés selon une séquence consensus et répertoriée. Les motifs d'acides aminés correspondent à ceux des centres [4Fe-4S], dont ils sont issus, modifiés soit par des insertions, soit par des délétions ou des substitutions de résidus.

Un centre [3Fe-4S] contient trois ions Fe<sup>3+</sup> ou deux Fe<sup>3+</sup> et un Fe<sup>2+</sup>. A l'état oxydé, les trois atomes de fer sont couplés antiferromagnétiquement via les atomes de soufre inorganiques ce qui aboutit à un spin total S=1/2 ; la protéine est donc paramagnétique et le centre présente un signal RPE centré à g=2. Après réduction à un électron, le centre [3Fe-4S] est constitué d'un ion Fe<sup>3+</sup> et d'une paire d'atome de fer à valence mixte Fe<sup>2.5+</sup> [14]. Ainsi le centre de spin total S=2 est dans un état paramagnétique uniquement détectable en mode parallèle en RPE.

Les centres [3Fe-4S] dérivent le plus souvent de la dégradation oxydative d'un centre [4Fe-4S] conduisant à l'inactivation de la protéine. Néanmoins, la présence de centre [3Fe-4S] dans certaines métalloenzymes actives a permis de montrer qu'ils pouvaient être des composants fonctionnels des protéines, dans les conditions physiologiques. Il faut aussi retenir la possibilité d'interconversion entre les formes [3Fe-4S] et [4Fe-4S], mécanisme proposé comme moyen de contrôle de certains métabolismes dans les réactions catalysées par certains centres [4Fe-4S], comme dans le cas de la Fumarate Nitrate Réductase (FNR) [15, 16].



### I.1.5-Les centres complexes

De nombreuses équipes cherchent à comprendre et à déterminer le fonctionnement des centres [Fe-S] issus de complexes enzymatiques atypiques qu'on ne fera que citer dans ce paragraphe. Ces centres métalliques qui présentent des caractéristiques communes ont intéressé nombres de chercheurs pour les raisons suivantes :

- Tout d'abord d'un point de vue fondamental, car ils sont impliqués dans des réactions rédox exceptionnelles d'activation de substrats très stables ( $H_2$ ,  $N_2$ ,  $CO_2$ ,  $SO_3^{2-}$ ).
- Ils sont directement impliqués dans l'activité enzymatique.
- Ils contiennent un nombre d'atomes de fer et de soufre supérieur à 4, et une spectroscopie RPE complexe. Ce type de centre, dont quelques exemples sont présentés sur la figure 4, est bien illustré dans des revues [3,16,<sup>17</sup>].

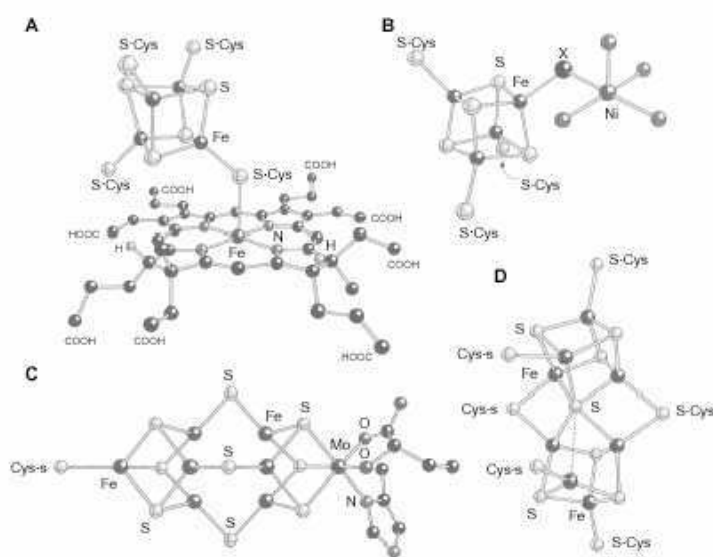


Figure 4 : **A** centre de la protéine Sulfite réductase.  
**B** centre de la protéine Carbone monoxyde déshydrogénase.  
**C** et **D** sont les centres de la protéine Fer-Molybdène de la nitrogénase d'*Azotobacter vinelandii*.

(Beinert H, Holm RH, Munck E (1997) Science 277:653-659)

## II-FONCTIONS DES PROTÉINES [Fe-S]

L'abondance des protéines fer-soufre dans un organisme donné est sujet à caution. En effet, une étude minutieuse de la littérature scientifique montre que 110 protéines [Fe-S] d'*Escherichia coli* ont été isolées jusqu'à présent. Toutefois, il a été montré qu'à ces 110 protéines, on pouvait ajouter 100 protéines potentielles [13]. Les protéines [Fe-S] représentent environ 5% de l'ensemble des protéines d'*Escherichia coli*. La potentialité d'un nombre important de nouveaux centres [Fe-S] à découvrir laisse entrevoir probablement de nouvelles fonctions biologiques et chimiques à ajouter à la liste sans cesse croissante du rôle des centres [Fe-S].

Si la première fonction biologique associée à des centres [Fe-S] fut celle de transférer des électrons, il n'en est plus de même aujourd'hui avec l'apparition de nouveaux types de centres [Fe-S]. En effet, à côté des centres "classiques" [2Fe-2S], [3Fe-4S], [4Fe-4S], il existe un certain nombre de centres exotiques (brièvement décrits ci-dessus). A cela s'ajoute une variété de ligands protéiques tels que des résidus histidine, arginine, aspartate, glutamate ou sérine ainsi que des exemples de coordination des centres par des ligands exogènes tels que la S-adénosylméthionine (SAM) dans le cas des protéines de la famille des radical-SAM, le Sirohème dans le cas de la Sulfite Réductase [3, 7]. Ces différences structurales induisent de nouvelles propriétés chimiques qui ont été exploitées par les enzymes, leur permettant ainsi d'avoir de nouvelles fonctions biologiques et catalytiques tels que la régulation de gène, la catalyse non-rédox et rédox, un rôle structural ou un rôle de préassemblage de centre [Fe-S] que l'on décrira dans la partie suivante.

### II.1-Les centres [Fe-S] dans le transfert d'électrons

Pendant de nombreuses années, le transfert d'électron fut la seule et unique fonction reconnue des centres fer-soufre. Les réactions de transfert d'électrons sont essentielles dans de nombreux processus biologiques incluant la photosynthèse, la respiration, et la fixation du diazote [7]. Les centres [Fe-S] servent de transporteur d'électron(s) soit au sein d'une même protéine, on parle alors de transfert intramoléculaire, soit entre deux ou plusieurs partenaires protéiques, on parle alors de transfert intermoléculaire.

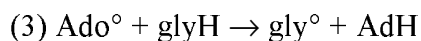
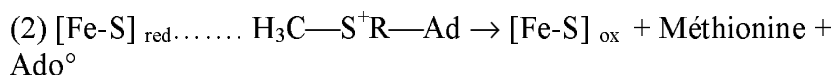
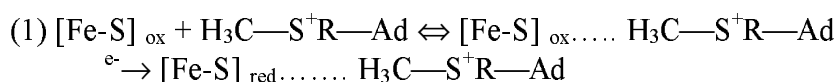
Parmi les protéines [Fe-S] possédant cette fonction on rencontre des protéines contenant des centres [1Fe], [2Fe-2S], [3Fe-4S] et/ou [4Fe-4S] et de centres complexes.

## II.2-L'activité enzymatique des centres [Fe-S]

Cette activité assurée par les centres [Fe-S] recouvre un très large éventail de fonctions biologiques qui sont :

- La catalyse non-rédox comme dans le cas des Déshydratases [<sup>18</sup>]. Dans ce cas le centre [Fe-S] est un [4Fe-4S] qui joue le rôle d'un acide de Lewis vis à vis du substrat.
- La régulation de gènes. En effet les centres [Fe-S] jouent un rôle clé pour la fonction de régulation de facteurs de transcription tels que les protéines FNR (Fumarate Nitrate Réductase), IscR, et SoxR d'*E. coli*. La FNR est un régulateur global qui contrôle l'expression de plus d'une centaine de gènes dans des conditions d'anaérobiose. La forme active, se liant à l'ADN, est un homodimère possédant un centre [4Fe-4S]<sup>2+</sup> par unité. En présence d'O<sub>2</sub>, il se produit une rapide conversion du centre [4Fe-4S]<sup>2+</sup> en un centre [2Fe-2S]<sup>2+</sup> qui rend inactive la protéine par perte de la dimérisation. SoxR joue un rôle de senseur dans le cadre du stress oxydant et il a été montré qu'elle réagissait avec le NO. Que ce soit dans le cas de stress lié au superoxyde ou à celui lié au NO, la protéine SoxR nécessite la présence d'un centre [2Fe-2S]<sup>2+</sup> pour être active. Dans ces conditions, la protéine est capable d'activer la protéine SoxS qui peut alors transcrire l'expression de protéines telle que la Superoxyde Dismutase. On peut aussi citer IscR répresseur de l'opéron *isc*. En effet, lorsque le pool de centres [Fe-S] dans la cellule et que les protéines [Fe-S] sont actives, IscR possède un centre métallique de type [2Fe-2S]. Sous cette forme, la protéine se lie à l'ADN et inhibe sa propre transcription ainsi que celle de l'opéron *isc* impliqué dans la biosynthèse des centres [Fe-S]. Pour terminer, on notera l'existence de senseur du fer, la protéine IRP (iron regulatory protein), qui contiennent un centre [4Fe-4S]. En cas de carence en fer, IRP est dépourvue de son centre métallique et peut alors, sous cette forme d'apoprotéine, activer et/ou inhiber la transcription de certains gènes [15,16].
- La catalyse rédox au cours de laquelle les centres [Fe-S] jouent un rôle de catalyseur couplé à un processus rédox. Ces types de centres sont représentés par les protéines de la super-famille des radical-SAM particulièrement étudiée au laboratoire, parmi lesquelles on trouve :

la Biotine synthase, la Lipoate synthase, la Ribonucléotide Réductase (RNR) mais aussi la protéine Thiomethyltransférase (MiaB) impliquée dans l'incorporation d'un groupement thiométhyle dans des ARN de transfert [19]. Toutes ces protéines possèdent un centres [4Fe-4S] typique avec un des atomes de fer différencié, sur lequel vient se fixer la S-adénosylméthionine. Le centre [Fe-S] réduit conduit en combinaison avec la SAM à la formation d'un radical, élément clé de la catalyse, comme dans le cas de la RNR (Figure 5).



$\text{H}_3\text{C}-\text{SR}-\text{Ad} = \text{AdoMet}$   
 $\text{Ado}^\circ = \text{radical } 5'\text{-déoxyadénosyle}$   
 $\text{AdH} = 5'\text{-déoxyadénosyle}$   
 $\text{gly}^\circ = \text{radical glycinyle}$

- (1) interaction de AdoMet avec le centre et réduction du centre [Fe-S]
- (2) réduction et clivage de Adomet.
- (3) génération d'un radical glycil.

Figure 5 : Rôle du centre [4Fe-4S] dans l'activation de la RNR anaérobie d'E. coli

(Fontecave, M., Mulliez, E. & Logan, D. T. in *Progress in Nucleic Acid Research and Molecular Biology*, Vol 72 95-127 (2002).)

Quoiqu'il en soit de l'état des études en cours sur la compréhension du fonctionnement de tout ces systèmes biologiques impliquant des centres [Fe-S], on comprend bien toute l'importance qu'ils représentent dans le monde du vivant et le rôle fondamental qu'ils peuvent jouer. C'est pourquoi depuis quelques années de nombreuses équipes de recherche, et en particulier notre laboratoire, s'intéressent à une problématique en amont, à savoir la biosynthèse des centres [Fe-S]. Il s'agit de comprendre comment les atomes de fer et de soufre sont mobilisés, puis assemblés, et incorporés enfin à la protéine [Fe-S] fonctionnelle. Comme nous le verrons dans la partie suivante de ce chapitre, ce processus implique une machinerie complexe de protéines dont l'une d'elles, la protéine SufA, fait l'objet de l'étude de cette thèse.

### III-BIOSYNTHESE DES CENTRES [Fe-S]

#### III.1-INTRODUCTION

Convaincu du rôle primordial que peuvent jouer les protéines [Fe-S] dans tous les organismes vivants, procaryotes et eucaryotes [3], il semblait légitime de s'intéresser au mécanisme de leur biosynthèse au sein de la cellule. Cette énigme a été le fruit d'intenses recherches durant ces dix dernières années, depuis la découverte par D. Dean et ses collaborateurs de protéines essentielles à l'assemblage des centres [Fe-S] de la Nitrogénase d'*Azotobacter vinelandii* [20] ; les protéines NifS et NifU. Cette découverte, associée à l'idée généralement admise que ni le fer ni le soufre n'existent sous forme soluble libre dans le cytoplasme, rendait invraisemblable le fait que la synthèse des centres [Fe-S] soit effectuée dans la cellule par simple réaction de l'apoprotéine (protéine sans son centre métallique) avec des atomes de fer et de soufre libres. De toute évidence des systèmes protéiques devaient être impliqués dans le processus de biosynthèse avec comme principales fonctions :

- la mobilisation les atomes de fer et de soufre de leurs sources de stockage respectives.
- l'assemblage sous la forme de centre [Fe-S].
- le transfert à la protéine cible.

et toutes ces réactions devant être parfaitement contrôlées de sorte que ni le fer ni le soufre ne soient libérés durant le processus, ce qui serait toxique pour la cellule.

de fait ces systèmes ont été identifiés. La cellule contient en fait un nombre très limité de "machineries" de biosynthèse pour l'ensemble des différents centres [Fe-S] au nombre de deux essentiellement (SUF et ISC décrit dans les pages suivantes). A la vue de cette stratégie, hautement économique, on peut s'interroger sur la spécificité de ces systèmes. En clair, par quel mécanisme ces "machineries" décident-elles d'assembler tel ou tel type de centre?

Les cellules d'*Escherichia coli* disposent essentiellement de deux de ces "usines" ou systèmes :

- Découvert en 1998 [21], le premier système appelé ISC (Iron Sulfur Cluster) a été largement étudié [7, 22]. Il est à noter l'existence de protéines homologues chez les eucaryotes [23].

- Le deuxième système nommé SUF (SULFur mobilization), découvert plus récemment, est principalement présent chez les bactéries, même si des homologues ont été récemment découverts chez la plante notamment chez *Arabidopsis thaliana* [24, 25, 26, 27].

Les gènes correspondants, respectivement *isc* et *suf*, sont en général organisés en opérons (Figure 6). Alors que les protéines codées à partir de l'opéron *isc* semblent intervenir dans un processus général pour l'assemblage des centres [Fe-S], celles issues de l'opéron *suf* joueraient plutôt un rôle dans des conditions de stress [28, 29, 30, 31].

- Récemment, un troisième système nommé CSD (Cysteine Sulfinate Desulfinate) a été découvert chez *E.coli* [32]. Ce système est spécifiquement impliqué dans la maturation du centre [Fe-S] de la quinolinate synthase, une protéine [4Fe-4S] impliquée dans la biosynthèse du NAD [33]. Ce système est extrêmement simple, comparé aux systèmes ISC et SUF, puisqu'il ne se compose que de deux protéines. Il est possible qu'il fonctionne grâce à la contribution d'autres protéines des systèmes SUF et ISC ou de nature encore inconnue.

### opéron *ISC*



### opéron *CSD*



### opéron *SUF*

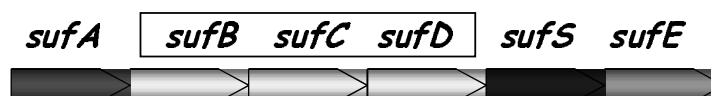
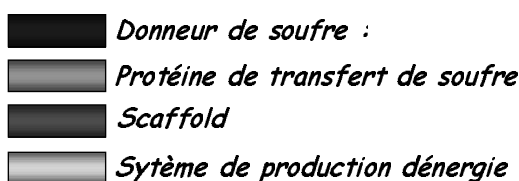


Figure 6 : Organisation des opérons ISC, CSD et SUF.

Code couleur :



### III.2-Le système SUF

Le système Suf d'*E. coli* est particulièrement étudié au laboratoire; il est composé de six protéines organisées en deux complexes SufSE (hétérodimère SufS+SufE) et SufBCD (hétéro trimère SufB+SufC+SufD), auxquels s'ajoute la protéine SufA. Il s'avère être le système est le plus simple parmi les opérons impliqués dans la biosynthèse des centres [Fe-S]. En effet chez les eucaryotes, on distingue au moins une vingtaine de protéines ISC impliquées dans ce processus.

#### III.2.1-Analyse physiologique

L'importance du système SUF dans la biosynthèse des centres [Fe-S] est venue des observations suivantes :

1) La mutation de *sufD* et *sufS* d'*E. coli* induit une instabilité du centre [2Fe-2S] de la réductase du fer ferrique (FhuF) d'*E. coli* [<sup>34</sup>].

2) Chez *E. coli* lorsque l'opéron entier *isc* est délété, 2 à 10% des activités des protéines [Fe-S] sont conservées. La surexpression de l'opéron *suf* restaure complètement l'activité de ces mêmes protéines, alors que l'inactivation des deux opérons, *isc* et *suf*, est létale pour la cellule [<sup>35</sup>].

3) La délétion de *sufC* chez *E. chrysanthemi* provoque une diminution de l'import de fer, ainsi que la diminution de l'activité d'enzymes à centres [Fe-S]. De plus on observe une plus grande sensibilité au stress oxydant [29].

L'importance du système SUF, dans les conditions de stress oxydant et de carence en fer, est venue des observations suivantes :

1) Le système SUF est régulé par le répresseur Fur, senseur de l'état de fer chez *E. coli* et *E. chrysanthemi* [<sup>36</sup>].

2) Il est aussi régulé par le régulon OxyR, senseur du peroxyde d'hydrogène chez *E. coli* [<sup>37</sup>].

3) L'expression des gènes *suf* est induite par l'addition de peroxyde d'hydrogène ou des générateurs de superoxyde ainsi que par la carence en fer [37,<sup>38</sup>].

### III.3-Les protéines Suf

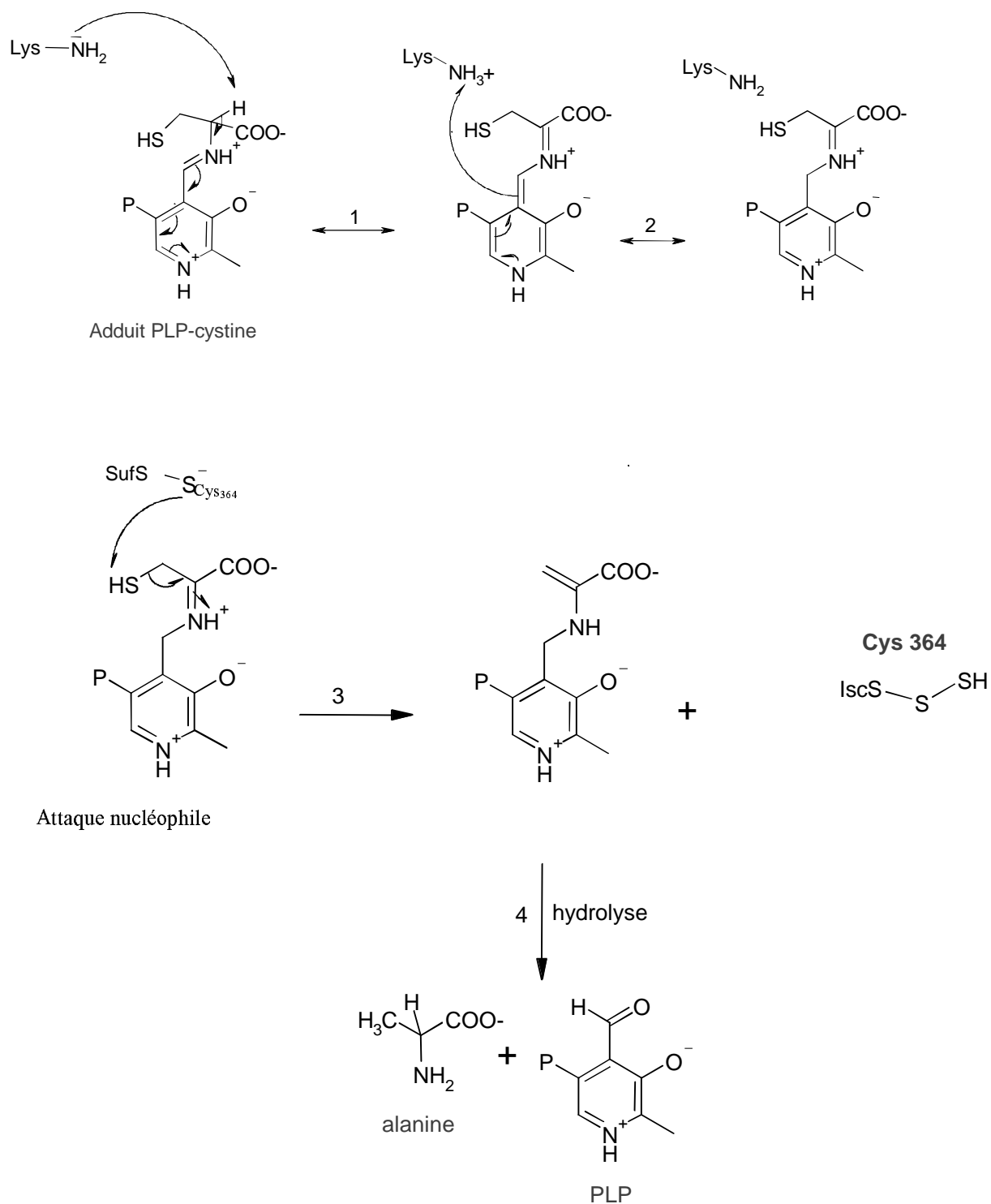
Il a été montré de manière non ambiguë que cinq des six protéines sont organisées en deux complexes distincts, nommés SufSE et SufBCD, tous deux localisés dans le cytoplasme. La protéine SufA, quant à elle, ne forme aucun complexe avec les autres protéines de l'opéron aussi bien *in vivo* qu' *in vitro* [<sup>39</sup>].

#### III.3.1-Le complexe SufSE

SufS est une protéine homodimérique possédant un cofacteur : le pyridoxal phosphate (PLP). Elle se caractérise par une activité cystéine désulfurase [<sup>40,41</sup>] en catalysant la conversion de la L-cystéine en L-alanine. Le site actif contient une cystéine conservée (Cys 364 chez *E. coli*), essentielle à l'activité.

SufE est une protéine homodimérique, qui forme un complexe 1:1 avec SufS. Elle ne possède pas d'activité cystéine désulfurase en soi. Associée à SufS, elle permet une stimulation de l'activité cystéine désulfurase de cette dernière [41,<sup>42</sup>]. Des études de mutagenèse dirigée ont montré que la Cys 51 de SufE était essentielle à cette stimulation de l'activité [<sup>43,44</sup>]. La réaction procède, premièrement, par la formation d'un persulfure sur la Cys 364 de SufS, par attaque nucléophile de cette cystéine sur un adduit PLP-cystéine (Figure 7 étape 3). Il s'ensuit la formation d'alanine et de PLP libre après l'hydrolyse. Dans un deuxième temps, un des atomes de soufre du persulfure de SufS est transféré à la Cys 51 de SufE par une réaction de transpersulfuration (Figure 8). Ce transfert spécifique d'atome de soufre a été parfaitement démontré par des expériences de marquage isotopique (<sup>35</sup>S) et par spectrométrie de masse [42,44]. L'utilisation de SufE mutée au niveau de la Cys 51 a démontré que cette cystéine est l'unique site accepteur de soufre.





**NB : P sur le cycle du PLP représente le phosphate**

Figure 7 : Mécanisme de l'activité cystéine désulfurase.

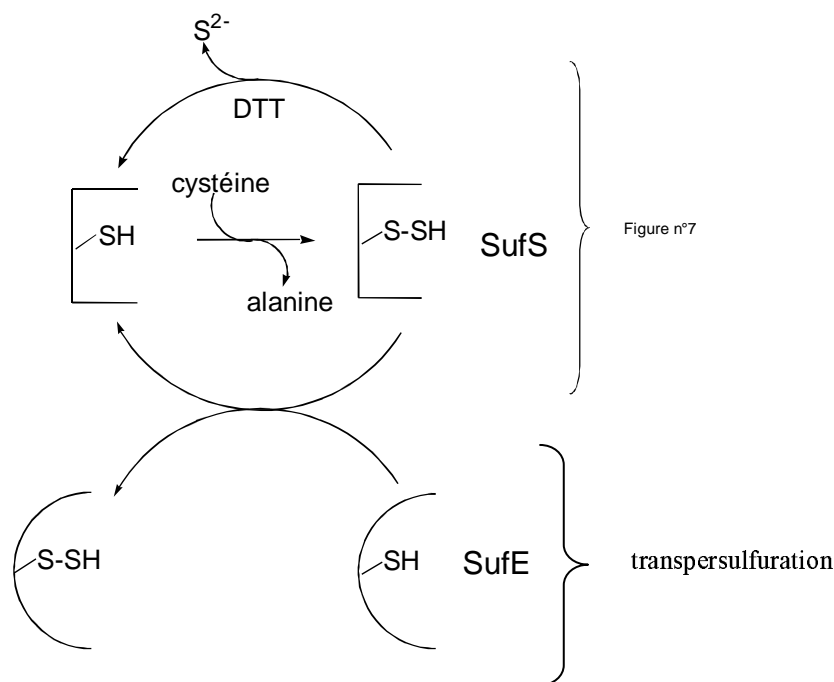


Figure 8 : Activité cystéine désulfurase et transpersulfuration.

(Fontecave M, Ollagnier de Choudens S, Py B, Barras F (2005) J Biol Inorg Chem:1-9)

Toutefois, le mécanisme moléculaire responsable de l'augmentation de l'activité cystéine désulfurase de SufS (associée avec SufE), n'est toujours pas connu à ce jour. La structure cristallographique de SufS d'*E. coli*, déterminée par Rayons X [<sup>45</sup>, <sup>46</sup>], montre que la cystéine 364 est éloignée du site où se trouve l'adduit PLP-cystéine, rendant la formation du persulfure quasi impossible. Ceci explique pourquoi l'activité cystéine désulfurase de SufS est extrêmement faible comparée aux autres cystéines désulfurases [<sup>47</sup>]. Ceci est cohérent avec de récentes études cinétiques montrant qu'en effet l'étape limitante de la réaction est l'attaque nucléophile de l'adduit PLP-cystéine par la cystéine 364 [<sup>48</sup>]. D'autre part, la structure cristallographique de SufE montre que la Cys51 est peu accessible. Elle est localisée sur l'extrémité d'une boucle qui lie deux brins  $\beta$  anti-parallèles et se situe dans une cavité hydrophobe. Par conséquent il est difficile d'envisager une réaction de transpersulfuration entre SufS et SufE sans invoquer un changement conformationnel des deux protéines, permettant le rapprochement entre la cystéine 364 et l'adduit PLP-cystéine d'une part et une meilleure accessibilité de la Cys 51 d'autre part. Ceci rendrait possible la réaction de persulfuration qui se traduirait par une augmentation de l'activité cystéine désulfurase de SufS. L'obtention de la structure tri-dimensionnelle du complexe SufS-SufE, non disponible actuellement, permettrait probablement de fournir quelques éclairages sur cette question.

### III.3.2-Le complexe SufBCD

Des expériences *in vivo* et *in vitro* ont montré que les trois protéines SufB, SufC et SufD d'*E. coli* et *E. chrysanthemi* forment un complexe stable : SufC présente une activité ATPase alors que les fonctions de SufB et SufD ne sont pas connues. Le complexe SufBCD augmente l'activité cystéine désulfurase du complexe SufSE [42], mais curieusement cette stimulation n'est pas ATP-dépendante. Des études récentes réalisées au laboratoire montrent que le complexe, et notamment SufBC, interagit avec SufE et, chose intéressante, un transfert de soufre de la protéine SufE à la protéine SufB a été démontré, avec formation de persulfures au niveau de résidus cystéines. En présence de fer et d'électrons, la forme soufrée de la protéine SufB est capable de former en son sein un centre [Fe-S], de type [4Fe-4S], réductible et stable, dont il reste à déterminer le rôle au sein du processus de biosynthèse des centres [Fe-S].

### III.3.3-La protéine SufA

SufA est une protéine cytoplasmique homodimérique, homologue à la protéine IscA, présent dans l'opéron *isc* (Figure 9). De nombreuses études ont été réalisées ces dernières années sur ces deux protéines [49, 50]. Les propriétés de SufA sont les suivantes : (i) SufA est obtenue principalement sous la forme apo après purification aérobie ou anaérobie [51] ; (ii) SufA est capable de chélater un centre [Fe-S] après traitement de l'apoprotéine avec du fer ( $\text{Fe}^{2+}$ ) et du soufre ( $\text{S}^{2-}$ ) en condition d'anaérobiose [51] ; (iii) Ces centres [Fe-S] sont instables et peuvent être transférés efficacement à une apoprotéine cible pour former des centres [2Fe-2S] ou [4Fe-4S] [51, 52] ; (iv) Le centre est très probablement chélaté par les trois cystéines conservées ( $\text{C}_{50}\text{X}_n\text{C}_{114}\text{GC}_{116}$  pour *E. coli*) dont l'espacement est caractéristique des protéines de type A. Ceci a permis de définir SufA comme une protéine d'échaffaudage ("scaffold") au sein de laquelle, le fer et le soufre sont préassemblés en centre [Fe-S] qui est ensuite transféré à différentes apoprotéines cibles.

Erc. SufA	-MQTHDVETFSIDEQAWQG-----ITLTDSAAVAHISNIMGRDESVOGIRLAVKQSGCAG	53
Ec. SufA	-MDMHS-GTFNPQDFAWQG-----LTLTPAAAIHIREIDAKQPGMVGVRLGVKQTGCAG	52
St. SufA	-MEIHS-DTFNPEDFPWQG-----LTLTPAAAAHIREIAEKQPGMLGVRLSVKQTGCAG	52
Ec. IscA	-MS-----ITLSDSAAARVNTFLANRKGFGIRLGVRTSGCSG	37
Av. IscA	-MA-----VTMTAAARHIRRSLDGRKGEGIRLGVRTSGCSG	37
Pa. IscA	-MA-----ISMTEAAAKHVQRSLGRKGEGIRLGVRTTGCAG	37
Erc. SufA	FGYVLELVHQDPDSRLYQREGARLFIPLKAMPISDGTVEVDFVQEGINQVFKFNNPKARH	113
Ec. SufA	FGYVLDVSVSEPDKDDLFEHDKAKLFVPLQAMPFIDGTVEVDFVREGLNQIFKFHNPKAQN	112
St. SufA	FGYVLDTVREPKDKDLVFEAEGAKLFAPLQAMPFIDGTVEVDYVQEGINQIFKFHNPKAQN	112
Ec. IscA	MAYVLEFVDEPTPEIDIVFEDKGKVVMDGKSLQFLDGTQLDFVKEGLNEGFKFTNPNVKD	97
Av. IscA	LAYVLEFVDEVASDQVFESHGKVIIDPKSLVYLDGTQLDFVREGLNEGFKFNNPNVRG	97
Pa. IscA	LAYVLEFVDELAEDLVFESHGKVIIDPKSLVYLDGTQLDFVREGLNEGFKFNNPNVRG	97
Erc. SufA	ACGGGSGFGL-----	123
Ec. SufA	ECGGGESFGV-----	122
St. SufA	ECGGGESFGV-----	122
Ec. IscA	ECGGGESFHV-----	107
Av. IscA	ECGGGESFNI-----	107
Pa. IscA	ECGGGESFNV-----	107

Figure 9 : Alignement de séquences de protéines homologues de type A

En général, après reconstitution, un monomère de SufA ne peut pas lier plus de deux atomes de fer et deux atomes de soufre. Cependant l'analyse de ces préparations par spectrométrie Mössbauer a montré que deux formes de centres, [2Fe-2S] et [4Fe-4S], existaient dans ces conditions [51]. SufA contiendrait un mélange de polypeptides chélatant un centre [2Fe-2S] par monomère et des polypeptides chélatant un centre [4Fe-4S]. Ce dernier pourrait se situer entre deux sous-unités ou être présent dans une seule unité, la deuxième étant dépourvue de centre. La présence de ces différents centres dans SufA peut s'expliquer de la façon suivante:

- SufA contient un centre [2Fe-2S] à partir duquel un centre [4Fe-4S] peut se former artificiellement à cause de l'excès de fer et de soufre présents dans le milieu réactionnel.
- SufA contient un centre [4Fe-4S] pouvant se décomposer en un centre [2Fe-2S].
- SufA est capable d'assembler les deux types de centres.

Aujourd'hui les questions essentielles qui se posent sont :

- quels types de centres sont réellement formés au sein de SufA? Déterminer la nature du centre [Fe-S] est un véritable casse tête. En effet, une hétérogénéité similaire de centres a été rapportée dans le cas des protéines IscA d'*A. vinelandii* et d'*E. coli* [49,<sup>53</sup>] alors que, seuls des centres [2Fe-2S] ont été détectés dans IscA de *S. pombe* et de *Synechocystis* [50,<sup>54</sup>]. De

plus, dans ce dernier cas les avis divergent quant à la quantité de fer par chaîne polypeptidique : soit un fer / soit deux fer par chaîne polypeptidique.

- quels sont les ligands des centres [Fe-S]? Les protéines d'échaffaudage possèdent trois résidus cystéines conservés, proposés comme étant les ligands des centres [Fe-S] [50, 53, 54, <sup>55</sup>, <sup>56</sup>, <sup>57</sup>], même si leur rôle exact n'a pas été démontré. En effet, des expériences menées *in vivo* sur Isa1p de la levure *S. pombe* ont montré que ces trois cystéines étaient essentielles à la fonction de la protéine [56,57]. Par contre, les mutants Cys→Ala de Isa1p sont tout aussi capables que la forme sauvage *in vitro* d'incorporer des centres [2Fe-2S] avec une efficacité de transfert identique [<sup>58</sup>]. Enfin, des expériences effectuées *in vitro* sur IscA de *Synechocystis* ont montré que seules les deux cystéines de la partie C-terminale étaient essentielles à la formation et au transfert du centre [2Fe-2S] [57,58]. On voit bien toute la complexité du problème. Une solution serait d'analyser la protéine directement dans la cellule par spectroscopie Mössbauer, en suivant la stratégie développée pour caractériser les centres [Fe-S] de la biotine synthase ou la protéine FNR *in vivo* [<sup>59</sup>,<sup>60</sup>]. On pourrait envisager de purifier SufA en anaérobiose mais l'instabilité du centre [Fe-S] rend l'expérience délicate.

L'obtention d'une structure tridimensionnelle de la forme holo (protéine possédant son centre métallique) de SufA serait extrêmement précieuse, pour l'identification des ligands du centre [Fe-S].

Un des objectifs de cette thèse est donc de déterminer la structure par RMN de la protéine SufA : la forme apo dans un premier temps, la protéine holoSufA dans un deuxième temps. Nous espérons ainsi pouvoir observer des modifications structurales au niveau des résidus cystéines et par conséquent déterminer les ligands du centre [Fe-S].

Il est toutefois important de mentionner ici que très récemment l'hypothèse qu'IscA et SufA sont des protéines d'échaffaudage a été remise en cause par des expériences biochimiques démontrant une forte affinité d'IscA d'*E. coli* pour le fer ( $K_d = 10^{-17}$  M) suggérant un rôle de chaperone à fer plutôt que de centre [Fe-S] [<sup>61</sup>,<sup>62</sup>]. SufA et aussi IscA de *A. Vinelandii* ont d'autre part été rapportées n'avoir aucune affinité pour le fer [49,53]. Ces résultats sont en contradiction avec ceux obtenus sur SufA et IscA d'*E. coli*. La pertinence physiologique de telles formes IscA-Fe ou SufA-Fe nécessite donc d'être considérée prudemment, à la vue de la récente structure de la protéine holo IscA, mais néanmoins ne doit pas être complètement écartée [<sup>63</sup>].

### III.4-Mécanisme d'assemblage des centres [Fe-S] et transfert

Peu de choses sont connues au niveau moléculaire en ce qui concerne des deux réactions dépendant de SufA à savoir l'assemblage du centre au sein de sa chaîne polypeptidique et le transfert du centre à des protéines cibles.

#### III.4.1-L'assemblage de centre

Le fait que le donneur de fer physiologique ne soit pas connu, rend cette étape difficile à étudier. Toutes les études *in vitro* utilisent un sel ferreux comme source de fer, et ainsi les conditions de réactions ne reproduisent pas exactement la situation *in vivo*. Par ailleurs, il est hautement probable que le donneur de soufre soit un persulfure localisé au niveau d'une cystéine réactive du système cystéine désulfurase. Dans le cas du système SUF, nous faisons l'hypothèse que cette cystéine est la cystéine 51 de SufE.

L'idée est de comprendre comment et dans quel ordre s'effectue l'insertion du fer et soufre au sein du site actif de SufA, défini comme comprenant les trois cystéines conservées (cys 50, cys 114, cys 116). Dans l'hypothèse de la formation d'un centre [2Fe-2S] à l'interface de deux sous-unités on peut proposer deux mécanismes [64]:

1) "Fe premier, S second" (Figure 10). Dans un premier temps, deux atomes de  $\text{Fe}^{2+}$  sont chélatés par les ligands cystéines conservés. Dans un deuxième temps, le soufre est incorporé via le système SufS. Cette étape implique une réduction à 2 électrons, du soufre ( $\text{S}^0$ ) en sulfure ( $\text{S}^{2-}$ ), fournis par les ions ferreux. La formation terminale du centre [Fe-S] requiert deux électrons supplémentaires qui pourraient provenir soit d'une source exogène soit d'un groupe actif dithiol.

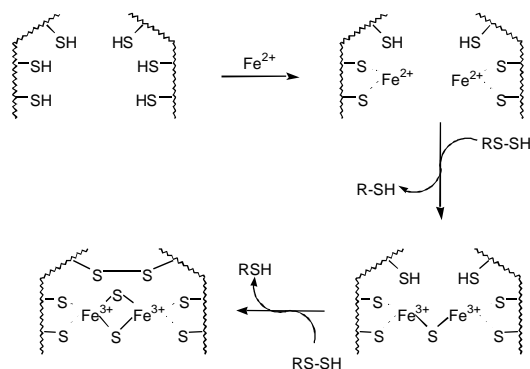


Figure 10 : Mécanisme dit "Fe premier, S second"

(Fontecave M, Ollagnier de Choudens S, Py B, Barras F (2005) J Biol Inorg Chem:1-9)

2) "S premier, Fe second" (Figure 11). Dans un premier temps, il y a réaction de transpersulfuration durant laquelle les cystéines nucléophiles de SufA acquièrent des atomes de soufre de SufE par l'attaque de ses persulfures, générant ainsi des persulfures sur SufA. Dans un deuxième temps, les ions ferreux sont incorporés. La formation du centre  $[\text{2Fe-2S}]$  à l'interface se produit après une réduction à deux électrons de chacun des deux atomes de soufre ( $\text{S}^0$ ), deux électrons fournis par les deux fer ferreux et deux par un dithiol rédox par exemple.

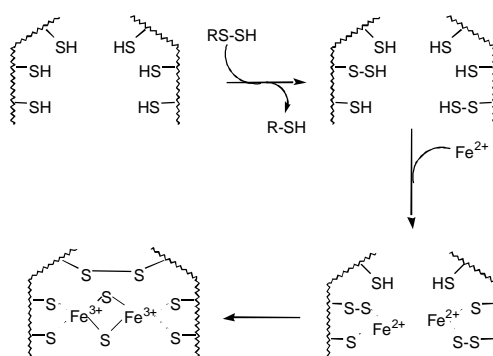


Figure 11 : Mécanisme dit "S premier, Fe second"

(Fontecave M, Ollagnier de Choudens S, Py B, Barras F (2005) J Biol Inorg Chem:1-9)

Des expériences biochimiques et spectroscopiques *in vitro* effectuées dans notre laboratoire (thèse Sendra Maïté, manuscrit soumis) ont montré que SufA accepte du soufre de SufE au niveau des trois cystéines conservées. Cette forme (SufA-SSH) en présence de fer et d'électrons est capable de former un centre [Fe-S].

Par ailleurs d'autres expériences biochimiques et spectroscopiques *in vitro* ont également montré que SufA est capable de lier le fer ferreux (1.6 fer par protéine) au niveau d'un site ne contenant pas les cystéines conservées. Même si ce fer est faiblement lié à la protéine (puisque un traitement à l'EDTA conduit à une protéine quasiment dépourvue de fer) on est capable, à partir de cette forme (SufA-Fe) en présence d'ions sulfures et d'électrons, de reconstituer un centre [Fe-S] au sein de SufA. Il semblerait donc que l'on soit en présence de deux sites sur la protéine ; un site où sont fixées les atomes de soufre au niveau des cystéines conservées et un site, non déterminé, où se fixerait le fer. Tous les résultats obtenus ne permettent pas de savoir qui, du soufre ou du fer, est inséré le premier.

Cette problématique de l'ordre d'insertion des atomes de fer et soufre dans les protéines d'échaffaudage a déjà été abordé avec la protéine IscU, mais il n'y a pas consensus sur le sujet [56, <sup>65</sup>]. Alors que certains proposent que IscU fixe d'abord le fer (avec une affinité égale à 68 $\mu$ M), d'autres proposent plutôt que le soufre vienne en premier [65]. Néanmoins, il n'a pas été possible de compléter l'assemblage du centre dans IscU par incubation de cette forme sulfurée avec du fer [56].

### III.4.2-Le transfert de centre

Que cherche t-on à connaître?

- le type de centre transféré
- la spécificité du transfert
- le mécanisme moléculaire

Des études menées au laboratoire, et effectuées avec la protéine holo-SufA et la biotine synthase (BioB) comme protéine acceptrice, ont permis de montrer que la réaction passe par la formation rapide d'un complexe entre les deux protéines. Aucun complexe ne se forme si la protéine cible contient déjà un centre [Fe-S], ou si l'apoSufA est incubée avec



l'apoBioB, ou l'holoSufA avec l'holoBioB [52], suggérant que le centre [Fe-S] contenu dans les protéines contrôle l'association. D'autre part, le centre [Fe-S] est protégé du solvant puisque la présence de chélateurs n'inhibe pas le processus de transfert [51]. Les derniers résultats du laboratoire montrent enfin que le transfert de centre est un processus rapide, prenant place en quelques minutes, et que l'étape lente est la dissociation du complexe ( $k = 0.03 \text{ min}^{-1}$ ) aboutissant à la forme active libre de holoBioB.

Lorsque les mêmes expériences ont été effectuées avec un mutant de la protéine BioB, dans laquelle une des cystéines ligand du centre a été changée en alanine, il a été observé que:

- le complexe est toujours présent, montrant que les cystéines ne participent probablement pas à la stabilisation du complexe [52].

- il n'y a pas de transfert du centre [Fe-S], suggérant que ce processus est guidé par l'attaque directe des cystéinates de la protéine acceptrice (BioB) sur le centre [Fe-S] de SufA. La réaction de transfert pourrait donc être décrite comme une attaque nucléophile au cours de laquelle les cystéines de la protéine SufA sont libérées et remplacées par celles de la protéine BioB (Figure 12).

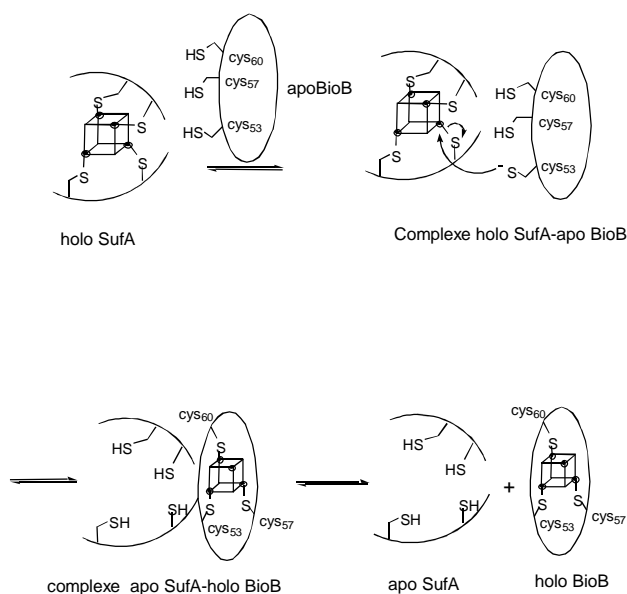


Figure 12 : Processus de transfert de l'holo SufA à l'apo BioB

(Fontecave M, Ollagnier de Choudens S, Py B, Barras F (2005) J Biol Inorg Chem: 1-9)

## Chapitre II

### Etude par RMN d'une protéine de grande taille

Si la détermination de la structure de la protéine SufA représente un enjeu important dans le cadre de nos recherches, notons qu'elle s'inscrit dans un processus plus général amorcé depuis quelques années dont l'objectif est la détermination des structures tridimensionnelles de l'ensemble des protéines identifiées. Ces entités, constituant entre 20 et 30 % de la masse cellulaire, possèdent des fonctions très diverses et essentielles à la vie de la cellule, et l'éthymologie de leur nom résume, à elle seule, leur importance. Après le décryptage complet du génome humain [66], qui permet d'identifier un nombre considérable de protéines par leur séquence d'acides aminés, les biologistes cherchent maintenant à déterminer leur structure tridimensionnelle et leur dynamique, espérant ainsi aider à la compréhension de la fonction de chacune d'entre elles. La vision directe de tels objets biologiques n'est malheureusement pas possible en raison de leur petite taille de quelques centaines d'Angströms (Å), alors que les microscopes optiques ne permettent d'observer que jusqu'à 400 nanomètres (nm). Les chercheurs ont développé trois techniques qui permettent d'atteindre des échelles de taille aussi petites. La première, la microscopie électronique, est idéale pour étudier des virus ou des complexes biomoléculaires puisqu'elle permet d'atteindre des résolutions d'environ 10 Å. Toutefois, la caractérisation des protéines requiert une résolution de l'ordre de l'Angström qui ne peut être atteinte que par les deux techniques que sont la radiocristallographie des rayons X et la Résonance Magnétique Nucléaire (RMN).

La plus ancienne, la radiocristallographie nécessite l'obtention d'un monocristal de protéine, sur lequel sont diffractés les rayons X. La déviation du faisceau informe sur la structure de la protéine. La RMN, qui n'a émergé qu'au cours des vingt dernières années, exploite les propriétés magnétiques des noyaux (spin) de certains atomes des molécules étudiées. Même si la radiocristallographie a permis la résolution de nombreuses structures, l'utilisation de la RMN peut se justifier sur trois plans : d'abord toutes les protéines ne cristallisent pas correctement et facilement ; ensuite, l'état cristallin peut induire des

modifications structurales de la protéine alors que la RMN permet l'étude de protéines en solution ; enfin la RMN permet d'accéder aux fluctuations structurales dans le temps par rapport à une conformation moyenne, donc à la dynamique de la protéine.

En plaçant la protéine dans un champ magnétique intense, il se produit une levée de la dégénérescence ou séparation des niveaux énergétiques des différents états de spin du noyau considéré. Cette séparation est propre à chaque noyau, car elle dépend de l'environnement chimique, et est proportionnelle au champ magnétique (Figure 1). La connaissance de cette séparation, permise par l'absorption puis l'émission de rayonnements électromagnétiques, rend donc possible l'identification de chaque noyau observable (étape d'attribution). On peut ensuite corréler des paires de noyaux entre eux pour obtenir des informations structurales mais aussi dynamiques. Cette technique est aussi très adaptée à d'autres applications telles que des études d'interactions intermoléculaires ou de repliements protéiques [67, 68]. Toutefois, la RMN est confrontée à ses propres contraintes, car la protéine étudiée doit être soluble et non agrégée à des concentrations de l'ordre du millimolaire mais, surtout, la masse de la protéine étudiée est une limitation majeure. Bien que des records de détermination de structures aient été enregistrés sur des monomères de 30 KDa [69, 70] et un homodimère de 47 KDa [71], la taille limite actuelle pour une étude de routine se situe aux alentours de 20 KDa. Au-delà, nous verrons que des phénomènes physiques viennent dégrader le signal RMN au point de ne plus pouvoir l'exploiter. Toutefois, cette limite est sans cesse repoussée grâce à diverses avancées techniques que nous décrivons.

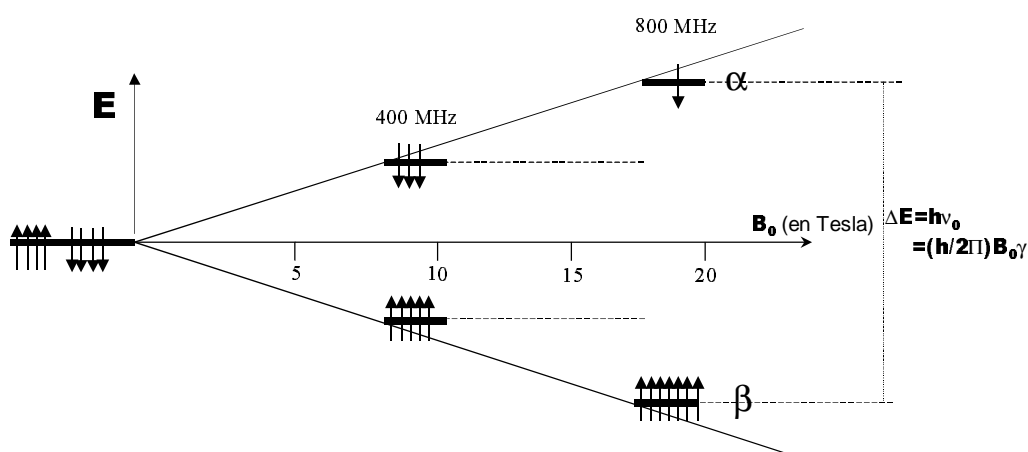


Figure 1: Principe de la séparation des niveaux d'énergie en RMN

Lorsqu'on place l'échantillon RMN dans le champ magnétique  $B_0$ , on observe une levée de dégénérescence des états de spins. L'écart en énergie est proportionnel à  $B_0$ . La différence de populations (qui obéissent à la loi de Boltzmann) entre les niveaux d'énergies permet le phénomène RMN.

---

## **I-ÉTUDE PAR RMN D'UNE PROTÉINE DE MASSE MOLÉCULAIRE ÉLEVÉE**

### **I.1-Limites imposées par la méthode de la RMN**

#### **I.1.1-Limites liées aux propriétés de l'échantillon biologique**

La sensibilité des expériences RMN croît principalement avec l'intensité du champ magnétique (en  $B_0^{3/2}$ ) et la quantité de spin identique (donc de molécules) présents dans l'échantillon en solution. Les volumes de détection restant quasiment inchangés quelque soit la sonde utilisée, pour un champ magnétique donné, le principal moyen d'accroître la sensibilité des expériences est donc d'augmenter la concentration de la protéine étudiée. Quelques dizaines de milligrammes de protéines (avec une pureté supérieure à 95%) suffisent à la préparation d'un échantillon de 500 mL à une concentration de l'ordre de 1 à 2 mM. La pureté a un impact décisif sur la qualité des spectres, mais aussi sur la longévité de l'échantillon qui peut se trouver dégradé par des protéases présentes. En revanche, l'expérience RMN ne cause pas intrinsèquement de dommages à la protéine. Cette dernière doit être stable pendant plusieurs semaines à une température pouvant se situer entre 10 et 40 °C.

#### **I.1.2-Limites liées à la taille de la protéine étudiée**

C'est la principale limite de la RMN comparée à la radiocristallographie (qui rencontrent aussi les mêmes problèmes décrits ci-dessus). Deux difficultés majeures apparaissent avec l'accroissement de la taille de la protéine.

D'une part, alors que la gamme de déplacements chimiques des différents atomes détectables reste inchangée (12 ppm pour le proton, 30 ppm pour l'azote et le carbone  $\alpha$ ), on observe des superpositions accidentelles des nombreux signaux ou pics RMN (Figure 2) qui peuvent être à l'origine de nombreuses ambiguïtés dans l'attribution des systèmes de spin.

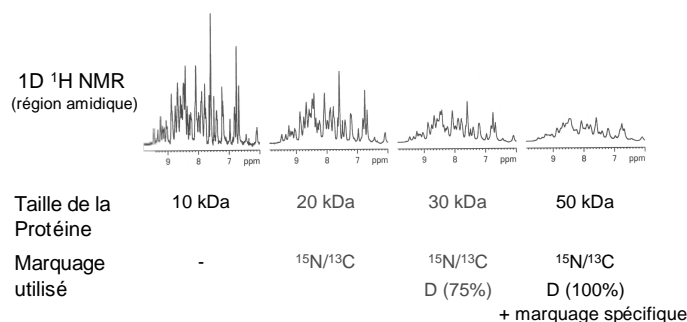


Figure 2 : RMN en fonction de la taille des protéines

D'autre part, la largeur à mi-hauteur des raies de résonances (des fonctions lorentziennes) est proportionnelle à la vitesse de relaxation transversale  $R_2$  :

$$S(\omega) = [1/R_2] / [1 + (1/R_2)^2 (\omega - \omega_0)^2] \quad (2.1)$$

où  $\omega_0$  est la fréquence de Larmor. Or  $R_2$  est proportionnel à  $\tau_c$  le temps caractéristique de réorientation globale de la protéine, lui même dépendant de la taille de la molécule. Dans l'approximation d'une molécule sphérique de rayon  $r$ ,  $\tau_c$  a pour expression :

$$\tau_c = 4\pi\eta r^3 / 3k_B T \quad (2.2)$$

où  $\eta$  est la viscosité du milieu,  $T$  la température et  $k_B$  la constante de Boltzman ( $1.3806 \cdot 10^{-23}$  J/K). Ainsi la largeur des signaux RMN augmente avec la taille de la protéine et donc que plus la taille de la protéine étudiée est grande, plus la sensibilité et la résolution des expériences RMN en sont affectées comme on peut le voir sur la figure 2.

## I.2-Les principales avancées visant à repousser les limites de la RMN

### I.2.1-Les spectromètres

La puissance des aimants, d'abord résistifs puis supraconducteurs, n'a cessé d'augmenter. En 40 ans, elle a été multipliée par 10, passant de 2.1 Tesla en 1962

(correspondant à une fréquence de résonance du  $^1\text{H}$  à 90 MHz) à 21 Tesla en 2002 (fréquence  $^1\text{H} \approx 900$  MHz). Cette évolution représente un progrès considérable puisque la résolution des spectres RMN augmente linéairement avec  $B_0$  et la sensibilité augmente en  $B_0^{3/2}$  [72]. Ainsi un spectre RMN issu d'un spectromètre 800 MHz est 2 fois mieux résolu et 6.4 fois plus sensible que celui issu d'un spectromètre 400 MHz. Ces améliorations entraînent une diminution sensible des recouvrements spectraux accidentels et une augmentation de l'intensité des raies de résonances. L'interprétation des données s'en voit simplifiée et les limites liées à la concentration de l'échantillon et à la taille de la molécule en sont repoussées. Parallèlement, la stabilité et l'homogénéité spatiale des bobines, deux paramètres importants pour la qualité des mesures, se sont améliorées.

### **I.2.2-Les cryosondes**

Une avancée technologique importante pour tenter d'abaisser la concentration minimale d'étude d'un échantillon est le développement des cryosondes. Celles-ci sont plongées dans un bain d'hélium, ce qui permet de minimiser le bruit thermique et d'augmenter ainsi la sensibilité des expériences d'un facteur 2 ou 3. Notons toutefois que cette amélioration de la sensibilité peut être perturbée par la présence de sel à forte concentration dans l'échantillon. En effet des courants électriques induits, aux niveaux de ces petites molécules ioniques, perturbent la sensibilité de la cryosonde et par conséquent amoindrissent le gain que cette technologie apporte.

### **I.2.3-Le marquage isotopique des échantillons pour une RMN multidimensionnelle**

La RMN des biomolécules est basée sur l'étude des interactions entre isotopes stables de spin nucléaires  $\frac{1}{2}$ . Parmi les atomes constituant les protéines, c'est-à-dire l'hydrogène, le carbone, l'azote, l'oxygène et le soufre, seul l'atome d'hydrogène possède une espèce majoritaire de spin  $\frac{1}{2}$ . L'azote et le carbone possèdent chacun un isotope de spin  $\frac{1}{2}$  ( $^{15}\text{N}$  et  $^{13}\text{C}$ ) mais dont la très faible abondance naturelle (respectivement 0.4% et 1.1%) ne permet pas d'atteindre une sensibilité suffisante pour les études structurales. Ceci explique en partie pourquoi la RMN des biomolécules a d'abord été une spectroscopie homonucléaire (1D) du

proton et ce jusqu'au début des années 1970. Toutefois, ces études ne pouvaient s'appliquer que sur de petits peptides pour les raisons que nous venons de décrire précédemment.

L'avènement de la RMN pulsée et l'apparition d'expériences RMN à deux dimensions (2D) au cours des années 1970, permettra de résoudre la première structure tridimensionnelle d'une protéine en 1985 [73]. Avec le développement de la spectroscopie multidimensionnelle (2D, 3D) hétéronucléaire (corrélation entre spins d'espèces atomiques différentes) vers la fin des années 1980, l'étape d'attribution des raies de résonances est plus aisée et la RMN, comme technique de détermination de structure 3D, devient accessible à des protéines de plus de 20 kDa. Les techniques de marquage isotopique deviennent par conséquent un outil essentiel pour l'étude en solution des biomolécules de cette taille.

Il existe deux types de marquage isotopique : le marquage uniforme et le marquage spécifique. Si l'étude structurale d'une protéine peut entièrement être réalisée à l'aide d'échantillon uniformément marqué, l'utilisation d'échantillons marqués spécifiquement peut compléter et même faciliter l'étude. En effet, comme seuls les résidus marqués sont détectables (un filtre hétéronucléaire est alors appliqué pour masquer les signaux des  $^1\text{H}$  non couplés), les spectres RMN de ce type d'échantillons se trouvent être nettement simplifiés. Les principaux inconvénients du marquage spécifique sont un coût élevé et la nécessité de produire plusieurs échantillons.

Un autre type de marquage a fait son apparition récemment. Il s'agit du marquage par intéine, c'est à dire un enrichissement isotopique sélectif d'une portion de protéine (de préférence définissant un module) [74]. Cette approche conduit à une simplification considérable des spectres RMN pour des protéines modulaires de grande taille par la réduction du nombre de résonances observables, tout en conservant la structure de la protéine intacte. Bien qu'encore peu pratiquée en raison de l'expertise qu'elle requiert, cette technique pourrait faire reculer la taille limite actuelle des protéines étudiables par RMN. Elle doit toutefois être associée à d'autres techniques pour remédier, comme nous le verrons, au problème de relaxation.

#### **I.2.4-La deutération**

La deutération, qui consiste à remplacer tous ou une partie des protons ( $^1\text{H}$ ) présents dans une protéine par des deutérons ( $^2\text{H}$  ou  $\text{D}$ ), est aussi un marquage isotopique pratiqué

pour l'étude des protéines en solution, bien qu'elle n'apporte pas de noyau supplémentaire à détecter (spin de  $^2\text{H}$  égal à 1).

La deutération aléatoire (uniforme) permet d'augmenter le rapport signal sur bruit des spectres hétéronucléaires multidimensionnels. Deux phénomènes sont à l'origine de cette amélioration : d'une part, la diminution de la vitesse de relaxation des spins  $^{13}\text{C}$  et  $^{15}\text{N}$  et, d'autre part, la suppression de la diffusion de spin.

En effet, la substitution d'un  $^1\text{H}$  par un D entraîne une diminution significative de l'interaction dipolaire entre cet isotope et un spin  $^{13}\text{C}$  (ou  $^{15}\text{N}$ ) directement lié, en raison de la différence entre les valeurs de rapports gyromagnétiques des deux isotopes ( $\gamma_{\text{H}} \approx 6.5 \gamma_{\text{D}}$ ). La vitesse de relaxation transversale, dont la principale source dans les expériences hétéronucléaires est l'interaction dipolaire, s'en trouve fortement diminuée [75]. La contribution des protons à l'augmentation de la largeur des raies de résonances est donc amoindrie.

La figure 3 représente le temps de relaxation transversale (spin-spin) des noyaux de la chaîne principale en fonction du  $\tau_c$  de la molécule, suivant que la protéine soit perdeutérée (c'est à dire que tous ses protons non échangeables ont été substitués par des deutérons) ou pas. On constate que l'amélioration est conséquente pour le  $\text{C}_\alpha$  ( $T_2$  divisé par un facteur 9) et le  $\text{H}_\text{N}$  ( $T_2$  divisé par un facteur 2). En revanche, aucun changement notable n'est observé pour le CO et le N puisqu'ils ne sont pas liés covalamment à des deutérons [76].

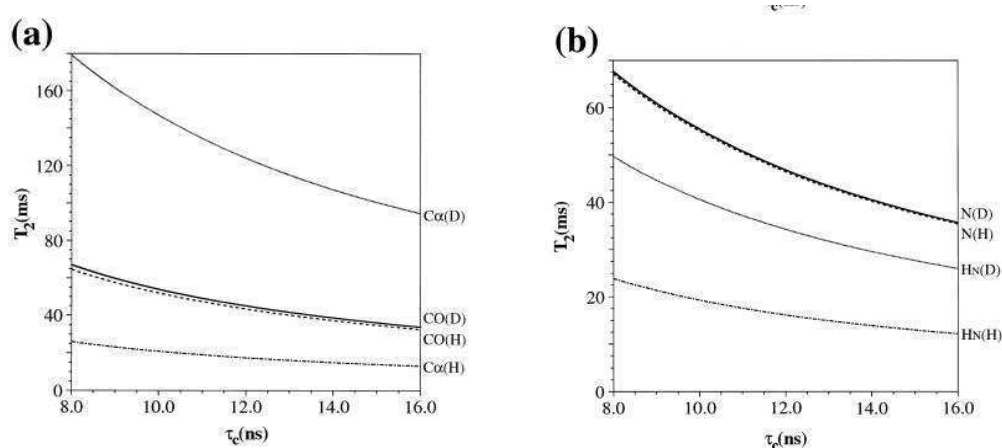


Figure 3 :  $T_2$  de différents noyaux calculés en fonction du  $\tau_c$  de la molécule et de la deutération [76]

De plus, la deutération entraîne une diminution de la densité spatiale des spins  $^1\text{H}$ , donc une réduction de la fuite de la magnétisation à travers de multiples chemins de relaxation



homonucléaire  $^1\text{H}$ - $^1\text{H}$  (Figure 4). Cet effet, appelé diffusion de spin, contribue à la relaxation longitudinale et augmente avec la masse de la protéine étudiée. Ainsi, en diminuant la diffusion de spin, on favorise la sélectivité de l'information dans le cadre des expériences hétéronucléaires (puisque l'on a des  $T_2$  plus long), améliorant ainsi la qualité des spectres.

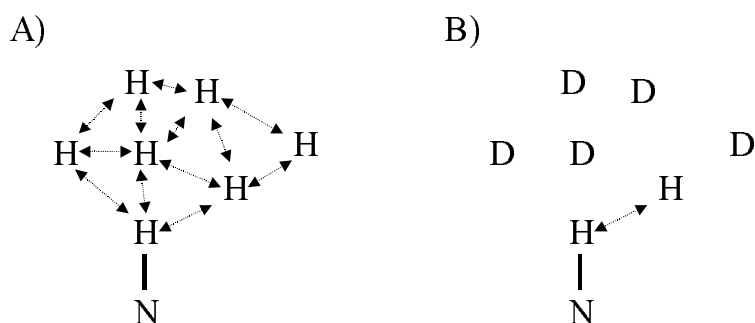


Figure 4 : Deuxième phénomène lié à la deutération :

- A) Effet de la diffusion de spin dans les protéines non deutérées.  
 B) Suppression de l'effet dans le cas de protéines deutérées.

Un taux de deutération compris entre 50 et 75 % suffit à compenser la diminution de la sensibilité, due à la perte du nombre de protons, par la diminution de la largeur des raies de résonances [77], même si les effets sont maximums sur un échantillon perdeuté. Notons que l'augmentation du rapport signal sur bruit des spectres RMN apportée par la deutération joue un rôle important dans la détermination de structure de grosse protéines par RMN [78].

Toutefois, les avantages remarquables qu'apporte la deutération ne sont pas sans contrepartie. Suite à la deutération, de la protéine d'intérêt, obtenue par surexpression dans un milieu complètement deutéré où la seule source d'hydrogène est l'eau lourde et le glucose. Il sera donc nécessaire d'effectuer l'échange des  $^2\text{H}_\text{N}$  par des  $^1\text{H}_\text{N}$ , indispensable à l'enregistrement des expériences RMN. Mais cet échange peut être plus ou moins long, particulièrement dans les régions de structures secondaires moins accessibles au solvant. D'autre part, le nombre de contraintes de distance, qu'il convient de mesurer pour le calcul de structure est fortement réduit par rapport à une protéine protonée. Il est d'ailleurs suggéré que pour des protéines de taille modérée, un taux de deutération de 50 % optimise la sensibilité des expériences pour l'attribution des chaînes latérales et représente un bon compromis avec l'enregistrement des spectres nOe [79]. Dans le cas de protéines de taille plus importante (plus de 25 kDa), des taux élevés de deutération sont alors nécessaires à l'enregistrement des

spectres bien résolus mais la collecte des nOe nécessaire à la détermination de la structure tridimensionnelle haute résolution devient difficile. Il s'agit donc de trouver un bon compromis de manière à pouvoir effectuer l'attribution de l'ensemble de la protéine et la collecte de contraintes nOe, ce qui peut se traduire par une période d'essais consommatrice en échantillons et temp. De nouvelles méthodes se développent actuellement tel que le marquage spécifique qui peut permettre d'obtenir un échantillon entièrement deutéré sauf au niveau des méthyles des valines, leucines et isoleucines [80]. Ces résidus étant très souvent impliqués dans le cœur hydrophobe des protéines, la mesure de contraintes nOe sur l'extrémité de leur chaîne latérale peut apporter des informations structurales très précieuses [81].

### 1.2.5-La spectroscopie TROSY

Les expériences qui corrélaient exclusivement les  $^1\text{H}_\text{N}$ ,  $^{15}\text{N}$  et  $^{13}\text{C}$ , ont donc été améliorées par la deutération, des  $^{13}\text{CH}_\text{n}$ , qui produit l'élimination de la relaxation dipolaire du  $^{13}\text{C}$  par le proton directement lié et qui permet la réduction de la largeur spectrale des  $^1\text{H}_\text{N}$  due au couplage dipôle-dipôle (DD) avec les protons dans l'espace environnant. Cependant, comme relaté avant, la deutération imposent certaines limites à la collecte d'informations structurales qui peuvent être obtenues par RMN.

C'est pourquoi, pour l'étude des protéines de grande taille, il a été développé l'expérience TROSY (Transverse Relaxation Optimised Spectroscopy) qui permet la sélection d'une seule des quatre corrélations issue du couplage  $J_\text{NH}$ . Ces corrélations résultent de l'interaction constructive ou destructive des deux phénomènes de relaxation transversale que sont le couplage DD et le CSA (chemical shift anisotropy) et qui ont un effet comparable à haut champ sur la relaxation [82]. Sur la Figure 5, on peut voir la différence d'information RMN entre une expérience HSQC (Heteronuclear Single Quantum Correlation) qui ne laisse apparaître qu'un seul pic à  $(\Omega_\text{N}, \Omega_\text{H})$  suite au découplage du  $^{15}\text{N}$  et du  $\text{H}^\text{N}$  et une expérience TROSY, dérivée de l'expérience HSQC.

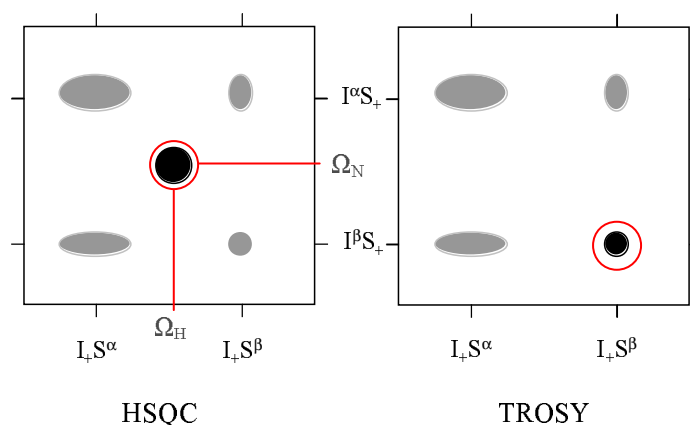


Figure 5 : Comparaison des largeurs de raies de résonances d'un spectre  $[^{15}\text{N}-^1\text{H}]$  HSQC et d'un spectre  $[^{15}\text{N}-^1\text{H}]$  TROSY  $[^83]$

Notons que ce phénomène n'est mesurable que sur les raies N-H amides (pour lesquels le maximum de cette contribution se produit à une fréquence de 1 GHz) et N-H aromatiques. D'autre part l'édition sélective de l'une des raies dans la séquence TROSY peut induire une perte d'intensité par rapport à une séquence HSQC. En revanche, sur un échantillon deutéré l'expérience TROSY apporte un gain en résolution et en sensibilité.

L'ensemble de ces développements permettent maintenant l'étude structurale de protéine dont la tailles se situe aux alentours de 25-30 kDA, même si, comme nous l'avons déjà cité, quelques records de résolution de structures ont été enregistrés sur des oligomères symétriques.

## II-ETAPES DE L'ÉTUDE PAR RMN DE PROTÉINE DE GRANDE TAILLE

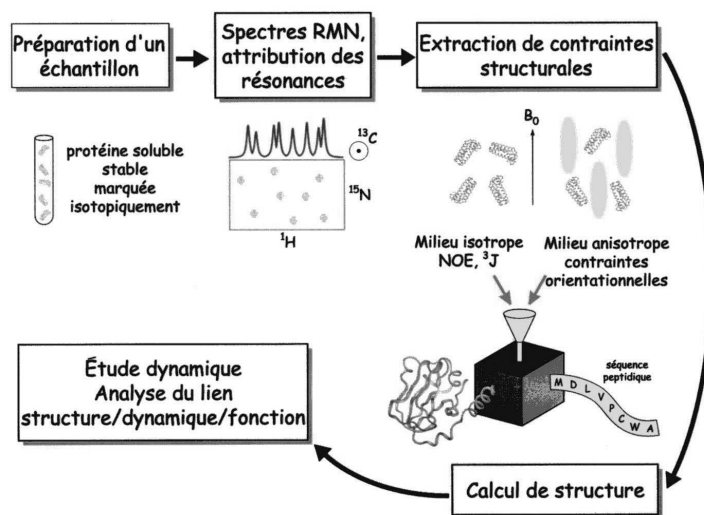


Figure 6 : Les différentes étapes d'une étude par RMN

L'approche expérimentale adoptée pour la caractérisation structurale et dynamique d'une protéine par RMN diffère suivant sa taille. La figure 6 schématise les différentes étapes de l'étude d'une protéine de grande taille à l'échelle RMN, c'est à dire de plus de 20 kDa enrichie en  $^{13}\text{C}$  et  $^{15}\text{N}$ .

### II.1-Expression de protéines recombinantes et marquées pour la RMN

La méthode la plus simple pour isoler la protéine étudiée est d'utiliser la cellule qui la produit naturellement. La RMN multidimensionnelle nécessite cependant des quantités importantes de matériel, quelques dizaines de milligrammes par échantillon, ce qui rend le travail de purification très lourd si la protéine n'est pas extrêmement abondante. La purification à partir de systèmes surproducteurs construits à partir du gène cloné est en général beaucoup plus aisée. Elle permet l'expression dans des micro-organismes (*E. coli*, levure) dont la croissance est rapide et les conditions de culture maîtrisées (division cellulaire toutes les 20 min à 37°C). De plus cette approche rend accessible le marquage isotopique ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$ ) uniforme ou sélectif. En effet l'utilisation d'organismes telle la bactérie *E. coli*,

facilement modifiée d'un point de vue génétique et adaptable, permet d'effectuer assez aisément les différents types de marquage. Cette étape de surexpression doit toutefois faire l'objet d'une évaluation minutieuse du meilleur système d'expression afin d'optimiser le rendement de production de la protéine et de minimiser son coût.

### II.1.1-Optimisation d'un système d'expression

Un système d'expression est optimal quand il permet la production d'une protéine recombinante pure, fonctionnelle et en quantité suffisante. Cela nécessite une adéquation totale entre le gène d'intérêt, l'organisme producteur, le vecteur d'expression, le milieu de production et les conditions de culture.

Le gène est une séquence d'ADN contenant l'information nécessaire à l'expression d'une protéine donnée. Il est obtenu à partir de l'ADN d'un organisme, puis après avoir été amplifié en quantité suffisante par PCR, il est introduit dans un vecteur d'expression ou plasmide. Ce dernier est un double brin d'ADN circulaire pouvant se répliquer de façon autonome dans l'organisme hôte et qui utilise la machinerie de la cellule hôte pour synthétiser la protéine d'intérêt.

Les systèmes d'expression existant aujourd'hui sont très nombreux et utilisent des stratégies diverses. Le plus souvent, *E. coli* est l'hôte de choix car c'est un organisme bien caractérisé, dont il existe de nombreux mutants (qui permettent d'amplifier ou inhiber une fonction biologique), et pour lequel un grand nombre de vecteurs, souvent commerciaux, ont été développés. On optimise la quantité de protéine produite en mettant le gène sous le contrôle d'un promoteur de transcription fort, c'est-à-dire capable d'induire la production de la protéine recombinante à une concentration de 10 à 30 % supérieure à celle de la totalité des protéines cellulaires. En assurant, d'autre part, un démarrage de traduction efficace pour éviter une fuite transcriptionnelle (expression de gène avant l'induction). En ce qui concerne le promoteur, on peut faire appel à système constitutif, c'est à dire non régulé, soit à un système inductible qui déclenche l'expression lorsqu'un stimulus extérieur est donné. Ces systèmes inductibles sont particulièrement intéressants pour exprimer des protéines toxiques pour

l'hôte. ; ils sont cependant délicats à utiliser car il faut déterminer, pour chaque cas, quel est le moment idéal pour déclencher l'expression.

Un autre aspect à prendre en considération est le choix du compartiment cellulaire dans lequel il est souhaité que la protéine soit exportée. Pour une sécrétion dans le cytoplasme ou dans le milieu extracellulaire, la présence d'une séquence signal sur le vecteur est requis en amont du gène d'intérêt.

Enfin, le plasmide peut contenir des « tags » d'affinité ou des protéines de fusion soit pour faciliter la purification (« tag » 6-histidines) ou la détection de protéine recombinante, soit pour prévenir la formation de corps d'inclusion en améliorant la stabilité et la solubilité (donc l'expression) des petites protéines (protéines Glutathion-S-transférase<sup>[84]</sup>, ubiquitine<sup>[85]</sup>). Sous condition de la présence de site de clivage spécifique, ces protéines de fusion peuvent être enlevées pendant les étapes de purification de la protéine d'intérêt.

### **II.1.2-Le milieu de production**

Afin de réduire les coûts de production, les protéines marquées sont produites en volume réduit de 1 ou 2 litres en général de milieu dit minimum dans lequel les sources en azote et/ou carbone sont contrôlées et marquées isotopiquement <sup>15</sup>N et/ou <sup>13</sup>C. Différents milieux ont été décrits dans la littérature qui vont du plus simple, le milieu M9 (comportant les éléments chimiques strictement nécessaires à la croissance bactérienne et dont la composition est donnée dans le chapitre matériels et méthodes)<sup>[86]</sup>, à d'autres milieux plus complets contenant des vitamines ou des oligo-éléments<sup>[87, 88]</sup>. Dans certains cas, il peut être nécessaire d'utiliser un milieu riche provenant d'algues marquées partiellement hydrolysées<sup>[89]</sup>.

### **II.1.3-Les techniques de marquage isotopique**

Le type de marquage isotopique dépend principalement de la taille de la protéine étudiée. En dessous de 10 kDa, la RMN du proton (expériences homonucléaires) suffit pour

l'étude de la protéine. Entre 10 kDa et 15 kDa, un simple marquage  $^{15}\text{N}$  permet d'améliorer la résolution spectrale. Entre 15 et 20 kDa, un double marquage  $^{15}\text{N}/^{13}\text{C}$  s'impose. Enfin, pour l'étude d'une protéine de plus de 20 kDa, un triple marquage est souvent requis.

D'un point de vue expérimental, le marquage isotopique d'une protéine chez *E. coli* est réalisé en utilisant du  $^{15}\text{N}$ -chlorure d'ammonium et/ou du  $^{13}\text{C}_6$ -glucose dans le milieu minimum M9. La deutération est effectuée en dissolvant les différents composants du milieu dans de l'eau lourde ( $\text{D}_2\text{O}$ ) stérile. L'eau lourde affectant négativement le métabolisme de la cellule (modification des longueurs des liaisons hydrogènes, des vitesses de réactions,...) [90], son incorporation entraîne donc une diminution de la croissance cellulaire qui peut affecter le rendement de production de la protéine d'intérêt.

Le marquage spécifique peut se pratiquer à partir d'acides aminés marqués disponibles commercialement. Un protocole a été décrit pour obtenir à moindre coût des protéines deutérées et marquées uniformément  $^{15}\text{N}/^{13}\text{C}$  et spécifiquement  $^1\text{H}$  sur les groupes méthyles des chaînes latérales [80].

### II.1.4-Préparation de l'échantillon RMN

La purification des protéines est un sujet vaste et dépasse largement le cadre de cette thèse. Elle s'effectue en plusieurs étapes qui vont de l'extraction de la protéine d'intérêt de l'organisme hôte (casse cellulaire, précipitation de l'ADN, centrifugation,...), à sa mise en solution et purification en utilisant différentes techniques chromatographiques (chromatographie d'exclusion, d'échange d'ion, d'interaction hydrophobe ou d'affinité) qui ont pour but d'isoler la protéine d'intérêt du reste des biomolécules qui sont restées en solution. Pour la RMN, le degré de pureté final n'est pas critique, une préparation homogène à 85-90 % sur gel de polyacrylamide SDS est probablement suffisante, à condition qu'il ne reste pas de contaminant majoritaire dépassant 3 à 4 %. Il est toutefois souhaitable d'avoir une pureté la plus élevée, afin de réduire les traces de protéases susceptibles de dégrader l'échantillon. La protéine est ensuite concentrée (par dialyse, ultrafiltration, chromatographie sur gel) en présence du tampon d'étude. Le tampon phosphate est souvent utilisé car, ne contenant pas de proton, il n'interfère pas avec l'enregistrement des spectres  $^1\text{H}$  des protéines non marquées. Actuellement des tampons tels que TRIS, HEPES, MES sont disponibles sous

forme perdeutérées. Il est d'autre part nécessaire d'optimiser les conditions de tampon et de pH pour répondre à celles imposées par la RMN tout en s'assurant que la protéine reste stable pendant plusieurs semaines.

La concentration minimale finale exigée pour une étude structurale est 0.5 mM dans un volume minimale de 450  $\mu\text{L}$  (300  $\mu\text{L}$  pour des tubes spéciaux). Afin d'éviter une dégradation de la protéine par des micro-organismes contaminants, la solution finale peut être stérilisée par filtration ou des faibles quantités d'antibiotiques et/ou d'azoture peuvent être ajoutées. Il est souvent nécessaire d'ajouter des inhibiteurs de protéases et enfin, le scellement de tube RMN sous un gaz inerte tel que l'argon, peut limiter les dommages oxydatifs causés dans le temps, par exemple sur les résidus méthionines et cystéines ainsi que tout développement bactérien.

## **II.2-Attribution des raies de résonances d'une protéine marquée**

Une fois obtenu un échantillon de protéine pur, concentré et stable, l'enregistrement des spectres RMN peut être effectué. Toutefois l'extraction des informations structurales et dynamiques n'est pas possible à ce stade de l'étude. Avant cela, il est impératif de connaître les déplacements chimiques de chacun des atomes de la protéine. Pour effectuer cette étape d'attribution des raies de résonances, trois types d'informations sont utilisées : les interactions à travers les liaisons (via les couplages scalaires  $^1\text{J}, ^2\text{J}$ ), les interactions à travers l'espace (via les couplages dipolaires) et l'environnement chimique (via les déplacements chimiques des atomes). La stratégie à employer pour l'attribution des raies de résonances va dépendre du type de marquage isotopique de l'échantillon. Dans le cas d'une protéine non marquée ou marquée uniquement à l'azote 15, la stratégie classique a été détaillée dans l'ouvrage de Wüthrich [91]. Dans le cas des protéines doublement ou triplement marquées, la méthode repose sur la combinaison d'expériences hétéronucléaires tridimensionnelles (3D) ou quadridimensionnelles (4D) qui corréleront 2 ou 3 types de noyaux différents par le moyen des couplages scalaires. Le principe du spectre 3D est schématisé sur la Figure 7 : L'idée est de lever l'ambiguïté au niveau de deux pics de corrélation dans un plan en ajoutant une troisième dimension qui devrait permettre de les séparer.



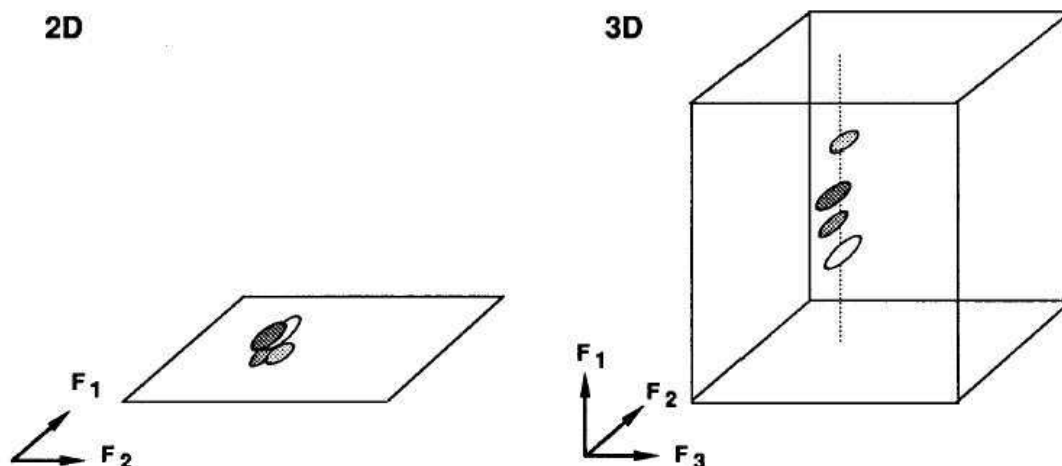


Figure 7 : Principe d'une expérience 3D

L'attribution de tous les noyaux observables par RMN d'une protéine doublement ou triplement marquée se fait en deux étapes : la première consiste en l'attribution de la chaîne principale et des carbones  $\beta$ , la seconde en l'attribution des chaînes latérales.

### II.2.1-Attribution de la chaîne principale et des $C_\beta$

Les expériences 3D triple résonance utilisées classiquement pour l'attribution de la chaîne polypeptidique d'une protéine doublement marquée sont toutes basées sur le même principe : la corrélation de l'azote et du proton d'un résidu  $i$  avec un troisième atome, un carbone dans la plupart des cas, du même résidu ou du résidu précédent [77,<sup>92</sup>]. Ces corrélations, dépendant du chemin de transfert de cohérence sélectionné, se font par le transfert successif de l'aimantation à travers 2, 3 voire 4 liaisons covalentes via des couplages scalaires  $^1J, ^2J$  relativement élevés (figure 8).

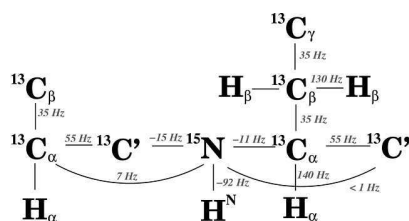


Figure 8 : valeurs des constantes de couplages  $^1J$  et  $^2J$  utilisées pour le transfert d'aimantation dans les protéines marquées  $^{15}N$  et  $^{13}C$ .

Ces expériences permettent d'identifier un carbone d'un résidu  $i$  et le même type de carbone issu du résidu  $i-1$  (en passant par le couplage scalaire  $N \leftrightarrow CO$ ). On obtiendra les déplacements chimiques du triplet ( $N_i$ ,  $H_i^N$ ,  $C_i$ ,  $C_{i-1}$ ). Les carbones peuvent être tour à tour les  $C_\alpha$ , les CO et les  $C_\beta$ . et les expériences correspondantes sont appelées HNCA et HN(CO)CA, HN(CA)CO et HNCO, HN(CA)CB et HN(COCA)CB. La notation entre parenthèses signifie que le transfert d'aimantation se fait via ces atomes mais qu'ils ne sont pas édités.

Une variante séquentielle est appliquée pour l'identification des fréquences en  $C_\beta$  dans les protéines doublement marquées. Contrairement aux échantillons deutérés, qui exploitent l'augmentation du temps de relaxation des spins des carbones liés à des deutérons en utilisant des expériences du type-HNC (HN(CA)CB et HN(COCA)CB) qui partent du proton amide, on préférera utiliser les expériences CB(CA)NH et CB(CACO)NH, qui sont alors plus sensibles, sur un échantillon doublement marqué mais protoné (Figure 9). L'ensemble des informations redondantes, permet de connecter des résidus successifs dans la séquence et donc de retracer la connexion de toute la chaîne peptidique.

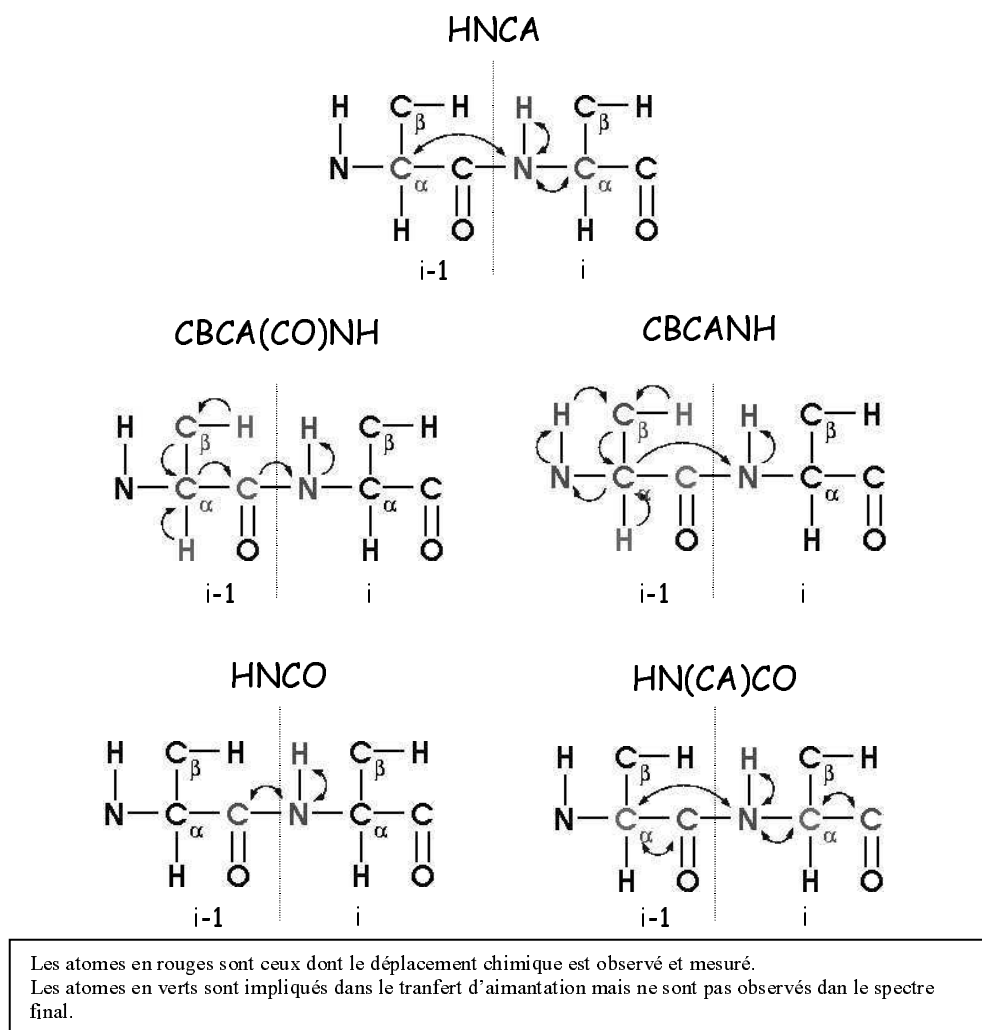


Figure 9 : Exemples d'expériences triple résonance pour l'attribution de la chaîne principale d'une protéine marquée  $^{15}\text{N}$  et  $^{13}\text{C}$ .

Enfin un quatrième couple d'expériences,  $\text{HN(COCA)H}$  et  $\text{HN(CA)H}$  permet toujours sur le même principe, d'attribuer les  $\text{H}_\alpha$  d'une protéine non deutérée.

Notons une différence pour les deux catégories d'expériences complémentaires : les expériences faisant intervenir le couplage scalaire  $\text{N} \leftrightarrow \text{CO}$  ne présente qu'un seul pic de corrélation par résidu alors que dans les expériences faisant intervenir le couplage  $\text{N} \leftrightarrow \text{CA}$ , peuvent se trouver deux pics par résidu, chacun provenant d'une corrélation avec le carbone  $i$  et  $i+1$ , en raison des valeurs de constantes de couplages  $^1\text{J}_{\text{N-C}\alpha}$  (-11 Hz) et  $^2\text{J}_{\text{N-C}\alpha}$  (7 Hz) proches. La différence d'intensité des deux pics, due au fait que  $|^1\text{J}_{\text{N-C}\alpha}| > |^2\text{J}_{\text{N-C}\alpha}|$ , peut

permettre de distinguer dans une même expérience le déplacement chimique provenant du  $C_i$  de celui provenant du  $C_{i-1}$  mais il est préférable de vérifier une telle attribution à l'aide de l'expérience complémentaire (qui sera dans ce cas l'expérience HN(CO)CA).

Dans le cas de très grosses protéines, l'identification d'un quatrième noyau peut s'avérer nécessaire. Des expériences en 4 dimensions (HNCACO et HNCOCA [93]) viennent encore améliorer la résolution des spectres. Enfin, il existe des expériences, dites à dimensionalité réduite qui présentent l'avantage de détecter 3 (ou 4) types de noyaux dans une seule expérience à 2 (ou 3) dimensions [94, 95, 96].

L'attribution de la chaîne principale s'effectue en exploitant l'information séquentielle contenue dans les triplets de fréquences combinée aux valeurs particulières des déplacements chimiques de certains résidus (Figure 10).

Cette étape d'attribution de la chaîne principale peut être assistée par des programmes d'attribution automatique tels que Autoassign [97] ou Alps [98] ou encore smartnotebook associé au logiciel Nmrview [99].

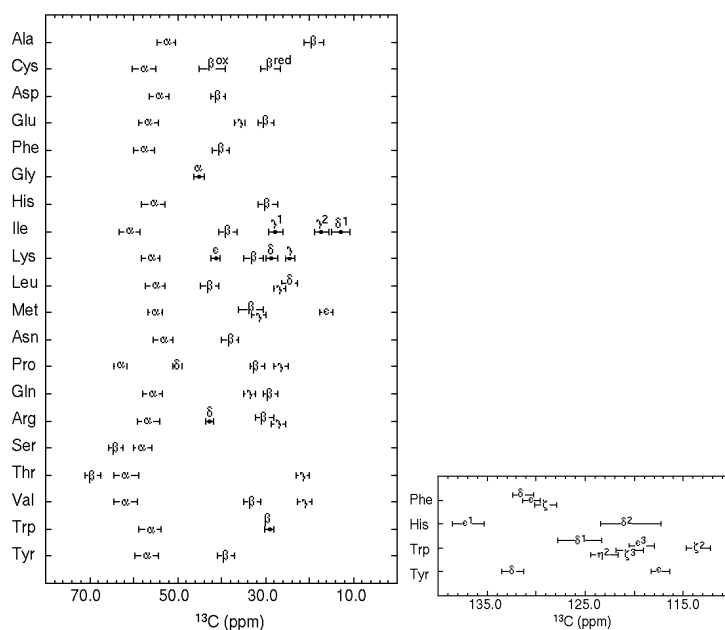


Figure 10 : Déplacements chimiques des  $^{13}\text{C}$  des chaînes latérales [100]

## II.2.2-Attribution des chaînes latérales

Les quatre expériences 3D les plus utilisées pour l'attribution des chaînes latérales des acides aminés d'une protéine doublement marquée  $^{13}\text{C}/^{15}\text{N}$  font intervenir un transfert  $^{13}\text{C}$ -TOCSY pour permettre le transfert de l'aimantation le long des chaînes.

Les deux premières sont de type 3D-HC(CACO)NH-TOCSY [ $^{101}, ^{102}, ^{103}$ ] et consistent, après excitation des protons aliphatiques, à relayer l'aimantation, via une étape de  $^{13}\text{C}$ -TOCSY, jusqu'au  $\text{C}\alpha$  puis à transférer à l'azote  $^{15}\text{N}$  et au proton  $\text{H}^{\text{N}}$  du résidu suivant. La séquence 3D-H(CCACO)HN-TOCSY permet d'éditer les fréquences  $^1\text{H}$  et on obtient des pics de corrélation ( $\text{H}_i^{\text{C}\alpha\text{C}\beta}$ ;  $\text{N}_{i+1}$ ;  $\text{H}_{i+1}^{\text{N}}$ ). La séquence 3D-(H)C(CACO)HN-TOCSY permet d'identifier les carbones de chaînes latérales et on obtient les pics de corrélation ( $\text{C}_i^{\alpha,\beta,\dots}$ ;  $\text{N}_{i+1}$ ;  $\text{H}_{i+1}^{\text{N}}$ ) (Figure 11).

Après avoir identifié les différents carbones et protons constituant la chaîne latérale d'un acide aminé, il reste à savoir quel proton est associé avec quel carbone. Deux expériences de type HCCH-TOCSY [ $^{104}, ^{105}$ ] permettent cette attribution. Dans un spectre 3D-(H)CCH-TOCSY on obtient des pics de corrélation entre les noyaux ( $\text{C}_i^{\alpha,\beta,\dots}$ ;  $\text{C}_i^{\alpha,\beta,\dots}$ ;  $\text{H}_i^{\text{C}\alpha\text{C}\beta,\dots}$ ) et dans un spectre 3D-H(C)CH-TOCSY les pics corrélés correspondent aux noyaux ( $\text{H}_i^{\text{C}\alpha\text{C}\beta,\dots}$ ;  $\text{C}_i^{\alpha,\beta,\dots}$ ;  $\text{H}_i^{\text{C}\alpha\text{C}\beta,\dots}$ ) (Figure 11).

Le nombre important de transfert d'aimantation le long des chaînes latérales dans ces types de séquences limite leurs applications à des protéines de masse inférieure à 20kDa. Ces expériences ne peuvent pas être appliquées aux protéines fortement deutérées qui ne possèdent pas une forte concentration en protons aliphatiques car le transfert de cohérence part du proton qui est détecté.

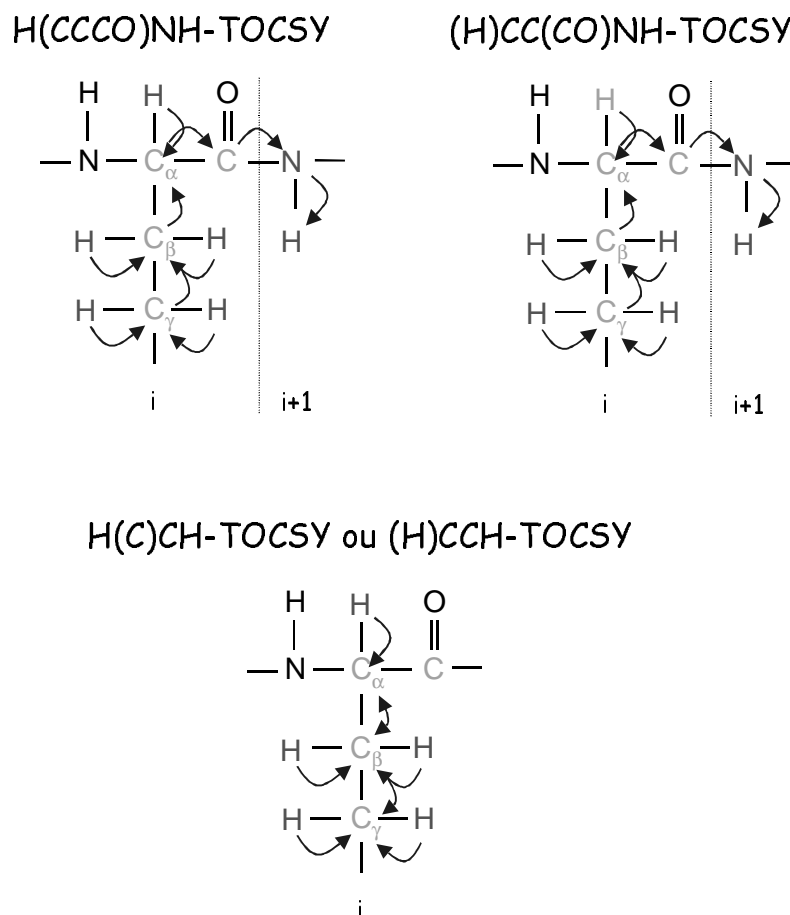


Figure 11 : Exemples d'expériences triple résonance pour l'attribution des chaînes latérales d'une protéine marquée  $^{15}N$  et  $^{13}C$ .

Toutefois des expériences ont été développées pour l'attribution des chaînes latérales de protéines partiellement deutérées [79]. Pour les protéines fortement deutérées, l'attribution des carbones des chaînes latérales peut se faire à l'aide de la séquence  $C(CACO)NH$ , équivalente à la séquence  $(H)C(CACO)NH$  précédemment décrite, mais au lieu de commencer par exciter les protons aliphatiques, ce sont les carbones aliphatiques qui sont polarisés au départ [106].

## II.3-Recueil des contraintes structurales

### II.3.1-Les contraintes mesurables classiques

L'étude structurale classique est basée sur la mesure d'un maximum de contraintes locales semi-quantitatives. Ces contraintes géométriques, des distances entre atomes et des angles dièdres entre liaisons, sont issues des trois observables RMN : la relaxation dipolaire, le couplage scalaire et le déplacement chimique.

#### -Contraintes de distances:nOe

La principale contrainte structurale utilisée en RMN des protéines est la distance  $r$  entre des paires de protons, proches dans l'espace (Figure 12).

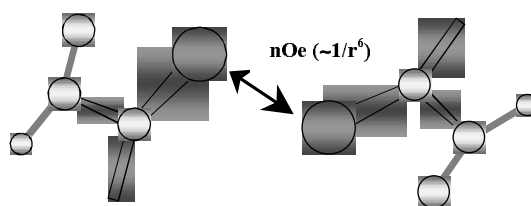


Figure 12 : Contrainte de distance mesurée par effet nOe

Cette information est mesurable grâce à la propriété de transfert d'aimantation longitudinale entre spins à travers l'espace, appelée la relaxation croisée ou l'effet Overhauser nucléaire (nOe = nuclear Overhauser effect). Cet effet nOe apparaît lorsque les deux noyaux sont assez proches dans l'espace ( $< 6\text{\AA}$  de distance) pour se transmettre de l'énergie par une interaction de type dipôle-dipôle. Cet effet de relaxation croisée est mesurable par RMN à l'aide de séquences de type NOESY [<sup>107</sup>]. La force de l'interaction nOe est déterminée par l'intégration du pic de corrélation résultant de l'échange d'aimantation entre deux spins en interaction dipolaire. Pour la mesure des volumes des corrélations nOe, on utilise un temps de mélange court afin d'être dans des conditions où le volume de chacune d'entre elles est directement relié à la distance intermoléculaire telle que:

$$I_{ij} \propto 1/r_{ij}^6$$

On néglige ainsi la diffusion de spin qui implique que le volume nOe soit une fonction complexe de plusieurs distances intermoléculaires. Théoriquement, si une distance interatomique  $r_{ref}$  est connue, alors en l'absence de mouvements internes locaux, les autres distances  $r_{ij}$  peuvent être déterminées par la relation :

$$r_{ij} = r_{ref} (I_{ref}/I_{ij})$$

où  $I_{ref}$  et  $I_{ij}$  sont les intégrales des pics de corrélation. En pratique, les temps de mélange sont toujours trop longs et les distances issues des nOe ne sont pas définies de manière exactes à cause de différents effets non désirables tels que la dynamique locale ou la diffusion de spin qui peuvent biaiser les mesures. Par conséquent, la relation entre volume nOe et distance intermoléculaire est inévitablement approximative. C'est pourquoi, on ne peut exprimer une distance dérivée d'un effet nOe que dans une gamme de distances permises plutôt que par une distance spécifique. Nous reviendrons sur cette problématique au moment de calcul de structure de SufA.

Notons qu'il est difficile d'attribuer tous les pics de corrélations nOe car ils peuvent correspondre à des résonances accidentellement dégénérées ainsi qu'à des résonances non attribuées, générant des ambiguïtés. Pour lever ces dernières, la stratégie classique est de calculer un jeu de structure préliminaire de basse résolution en utilisant uniquement les pics de corrélations nOe attribués de façon non ambiguë, et ensuite d'utiliser ces structures comme structure de base pour de nouveaux calculs qui devraient permettre de faciliter l'analyse et l'attribution des pics de corrélation nOe restants. Un certain nombre de processus d'attribution semi-automatique des contacts nOe ont été proposées (AutoStructure [<sup>108</sup>]) mais d'autres approches telle que ARIA [<sup>109</sup>], basées sur l'utilisation de ces contraintes ambiguës, se sont développées ces dernières années.

#### -Les angles dièdres:

Les angles dièdres, formés entre deux noyaux reliés par trois liaisons covalentes, constituent un deuxième type de contraintes qui définissent la conformation que peut prendre la chaîne polypeptidique. En effet, la chaîne principale d'une protéine est une suite de plans peptidiques qui s'articulent les uns par rapport aux autres au niveau de leurs liaisons N-C $_{\alpha}$  et C $_{\alpha}$ -C' (ou de leurs angles respectivement  $\phi$  et  $\psi$ ) (Figure 13). De même, l'orientation d'une



chaîne latérale d'un acide aminé peut être décrite par ses angles dièdres  $\chi_1, \chi_2, \dots$  définis au niveau des liaisons  $C_\alpha-C_\beta, C_\beta-C_\gamma, \dots$

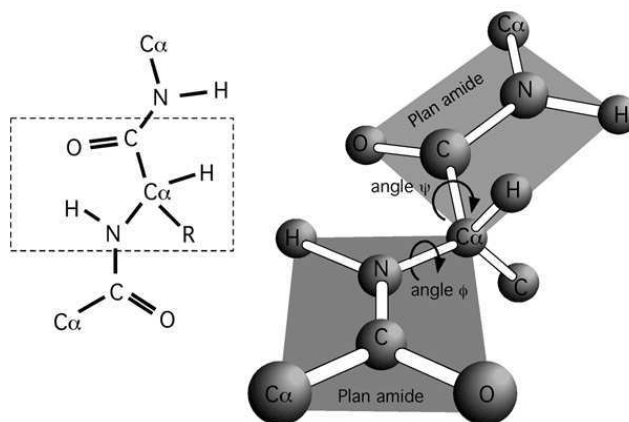


Figure 13 : Représentation des angles  $\Phi$  et  $\psi$  entre deux résidus

Le paramètre RMN qui contient l'information des angles dièdres d'une protéine est la constante de couplage scalaire homonucléaire ou hétéronucléaire  $^3J$ . Ce couplage entre deux noyaux séparés par trois liaisons covalentes est une fonction de l'angle dièdre  $\theta$  suivant l'équation de Karplus [<sup>110</sup>] :

$$^3J = A \cos^2 \theta + B \cos \theta + C$$

où A, B et C dépendent des noyaux impliqués dans les liaisons et ont été déterminés empiriquement. Les couplages les plus mesurés sont  $^3J_{H^N H^\alpha}$ ,  $^3J_{H^\alpha N}$  et  $^3J_{H^\alpha H^\beta}$  et dépendent respectivement des angles  $\phi$ ,  $\psi$  et  $\chi_1$ . Toutefois, l'allure des courbes de Karplus montre qu'à une constante mesurée peut correspondre jusqu'à quatre valeurs d'angles dièdres. Ce paramètre est donc dégénéré.

Cette information peut être complétée par des mesures de couplages scalaires inter-résidus  $^3J_{NCO}$  à travers les liaisons hydrogènes  $N-H \cdots O=C$ . En effet, les éléments de structure secondaire sont stabilisés par un réseau régulier de liaisons hydrogènes dont l'identification expérimentale par ce paramètre, bien qu'il ne soit pas toujours très facile à mesurer (expériences peu sensibles à cause de la faible valeur de la constante de couplage, environ 0.3 Hz, pouvant nécessiter des délais longs dans la séquence d'acquisition) confirme leur présence et permet de les contraindre pendant le calcul de structure. [<sup>111</sup>, <sup>112</sup>].

-Le déplacement chimique :

Enfin, le déplacement chimique des atomes de la chaîne principale et du début de la chaîne latérale est extrêmement sensible aux conformations locales [<sup>113</sup>]. A partir de ces valeurs, il est possible d'identifier avec une bonne précision les régions de structure secondaire d'une protéine, soit en calculant un index de déplacement chimique (CSI = chemical shift index) par rapport à une valeur standard par résidu mesurée dans des pelotes statistiques [<sup>114</sup>, <sup>115</sup>], soit en calculant, à partir d'une base de données, les angles dièdres  $\phi$  et  $\psi$  de la protéine à l'aide par exemple du programme TALOS [<sup>116</sup>]. Ce logiciel se base sur une prédiction empirique des angles de torsions  $\phi$  et  $\psi$  en utilisant les combinaisons des cinq types de déplacement chimique ( $H_\alpha$ ,  $C_\alpha$ ,  $C_\beta$ , CO, N) et les structures d'une vingtaine de protéines qui constituent sa base de données. En pratique, TALOS compare les déplacements chimiques et la séquence en acides aminés d'un triplet de résidus issus de la base de données à celle d'un triplet issu de la protéine. De cette comparaison, TALOS garde les 10 meilleurs résultats pour le résidu central de ce triplet. Ensuite, le programme donne en sortie les valeurs  $\phi$  et  $\psi$  du résidu central (en fait des valeurs moyennes et leurs déviations standards). Si les valeurs  $\phi$  et  $\psi$  sont en désaccord, TALOS ne fait pas de prédiction pour le résidu central.

### II.3.2-Les contraintes orientationnelles

Si l'approche classique de la RMN, basée essentiellement sur la détermination de contraintes locales, est très performante pour l'étude de protéines globulaires ou compactes, une information à plus longue portée peut parfois faire défaut. C'est notamment le cas lors d'études de protéines deutérées (qui ont peu de protons), multidomaines (pour lesquelles un jeu de distance inter-module est trop souvent limité pour pouvoir définir leurs positions relatives) ou de formes allongées (dont la géométrie spécifique ne permet pas de mesurer assez de nOe). Pour toutes ces molécules, une structure 3D haute résolution ne peut être obtenue à partir des seules contraintes de distances de type nOe. C'est pourquoi le développement d'une nouvelle approche, basée sur l'obtention de contraintes orientationnelles et à longue portée, a suscité un grand intérêt ces dernières années. Ces contraintes sont appelées orientationnelles parce qu'elles dépendent de l'orientation d'un vecteur de la molécule étudiée (par exemple le vecteur N-H<sub>N</sub>) par rapport au champ

magnétique statique. Chaque vecteur aura donc une orientation propre qu'on cherche à déterminer.

Pour notre part nous avons mesuré des couplages dipolaires résiduels (RDC residual dipolar coupling) dans un milieu orientant nous permettant d'obtenir un jeu de contraintes supplémentaires qui est injecté dans le calcul de structure au même titre que les autres contraintes. Ces contraintes orientationnelles nous donnent par contre une information locale, en terme de structure que les contraintes de distances de type nOe ne permettent pas d'obtenir. Par conséquent nous les avons utilisé pour l'affinement de la structure.

### II.3.2.1- Interaction dipolaire

#### II.3.2.1.1- Aspect théorique

Le spin nucléaire génère un champ magnétique autour de lui dans la direction de son moment magnétique qui peut interagir avec un second spin nucléaire qui possède lui-même son propre champ magnétique local. Le couplage dipolaire est l'interaction mutuelle de ces deux dipôles magnétiques qui pour deux spins I et S a un Hamiltonien de la forme (1), considérant que seule la composante longitudinale z (donc parallèle à  $B_0$ ) du champ magnétique local du spin  $I_i$  sur le spin  $I_j$  est significative

$$H_{ij}^D = - \gamma_I \gamma_S \mu_0 \hbar / 8\pi^2 I_z S_z [(3\cos^2\theta(t) - 1)/2r^3(t)] \quad (1)$$

$\gamma_i$   $\gamma_j$  sont les rapports gyromagnétiques des spins considérés, r leur distance relative et  $\theta$  l'angle entre vecteur r qui relie les deux spins, et le champ magnétique statique  $B_0$ . On observe donc un couplage  $D_{ij}$

$$D_{ij} = - \gamma_I \gamma_S \mu_0 \hbar / 8\pi^2 \langle 3\cos^2\theta(t)/2r^3(t) \rangle \quad (2)$$

Si on considère d'autre part que les couplages sont mesurés dans une molécule supposée rigide, alors les distances r restent constantes, et on peut exprimer  $D_{ij}$  de la manière suivante :

$$D_{IS} - D_{IS}^{max} / 2 \langle 3\cos^2\theta - 1 \rangle \text{ avec } D_{IS}^{max} = - \gamma_I \gamma_S \mu_0 \hbar / 8\pi^2 \quad (3)$$

Ainsi la valeur du couplage dipolaire dépend de l'orientation du vecteur  $r$  par rapport au champ statique, orientation pondérée par sa probabilité d'être dans l'orientation  $\theta$ . Dans un milieu isotrope, où toutes les orientations sont possibles et équiprobables, le couplage dipolaire résiduel est moyenné à 0. Par contre dans un milieu anisotrope tel que les cristaux liquides, l'orientation préférentielle entraîne une moyenne non nulle des termes de couplage dipolaire. Cette approche repose sur la description de l'orientation de la molécule par un tenseur d'alignement (de trace nulle) qui décrit la distribution des orientations des molécules dans le milieu anisotrope. Clore et *al* [<sup>117</sup>] ont montré que la constante de couplage  $D_{ij}$  peut être simplifiée en la relation suivante

$$D_{IS}(\theta, \varphi) = D_a^{IS} \{ (3\cos^2\theta - 1) + 3/2 R (\sin^2\theta \cos 2\varphi) \} \quad (4)$$

où  $D_a^{IS}$  et  $D_r^{IS}$  sont les composantes axiale et rhombique du tenseur  $\mathbf{D}$  données par :  $D_a^{IS} = 1/3[D_{zz}^{IS} - (D_{xx}^{IS} + D_{yy}^{IS})/2]$  et  $D_r^{IS} = 1/3[D_{xx}^{IS} - D_{yy}^{IS}]$ .  $R$  est le rhombicité définie par  $D_r^{IS} / D_a^{IS}$  et est toujours positive ;  $\theta$  est l'angle entre le vecteur interatomique  $r$  et l'axe  $z$  du tenseur ;  $\varphi$  est l'angle de la projection du vecteur interatomique  $r$  dans le plan  $(x,y)$  relativement à l'axe  $x$ .  $D_a^{ij}$  inclut différentes constantes tels les rapports gyromagnétiques des deux noyaux  $\gamma_I$  et  $\gamma_S$ , l'inverse de la distance entre les deux noyaux au cube  $\langle r^{-3}(t) \rangle$  moyennée dans le temps et le paramètre d'ordre général  $S$  pour les fluctuations angulaires du vecteur internucléaire. L'expression (4) montre que la valeur du couplage dipolaire va dépendre de l'orientation du vecteur interatomique  $r$  dans le référentiel du tenseur d'alignement  $\mathbf{D}$ . De ce fait, la valeur du couplage dipolaire permettra donc d'obtenir des informations concernant l'orientation des différents vecteurs de la protéine par rapport à ce référentiel. Cependant, le couplage dipolaire résiduel est dégénéré. La détermination de l'orientation de vecteur sera par conséquent obtenue à partir de la mesure du couplage dans au moins deux milieux orientant. L'alignement de la molécule dans le deuxième milieu, permet d'obtenir deux jeux de données avec des tenseurs d'alignements différents, qui vont permettre de lever la dégénérescence. Pour l'analyse de SufA, nous avons utilisé un seul milieu parce que nous voulions utiliser les RDC uniquement au cours du calcul de structure pour l'affinement, donc à partir d'une structure connue dont on a déterminé les composantes du tenseur.

Afin de pouvoir utiliser les RDC pour l'affinement de la structure de la protéine, les composantes axiale et rhombique du tenseur d'alignement doivent être connus. La méthode la plus efficace pour déterminer les éléments du tenseur d'alignement est la décomposition en valeurs singulières [<sup>118</sup>]. Il y a cinq éléments indépendants dans le tenseur d'alignement. Ils

correspondent aux composantes axiale et rhombique et son orientation, définie par les trois angles d'Euler qui caractérisent la rotation du repère du tenseur d'alignement par rapport au champ magnétique statique. Pour déterminer ce tenseur, il est nécessaire que le nombre de RDC soit plus grand que le nombre d'éléments à calculer. C'est le cas d'une protéine où ce nombre est bien supérieur à cinq.

Le programme MODULE [<sup>119</sup>] développé au laboratoire RMN de l'IBS permet de déterminer le tenseur d'alignement via les RDC mesurés pour des molécules partiellement orientées en optimisant les paramètres du tenseur par une régression non linéaire utilisant un recuit simulé.

Clore et *al* [117] propose une approche qui consiste à se servir de l'histogramme de l'ensemble des couplages dipolaires résiduels normalisés. Si la distribution des vecteurs internucléaires est uniforme, un tel histogramme ressemble à un spectre de poudre observé pour l'anisotropie du déplacement chimique en RMN à l'état solide. En effet, dans un histogramme composé de plusieurs types de RDCs (<sup>1</sup>D<sub>NH</sub>, <sup>1</sup>D<sub>CON</sub>, <sup>2</sup>D<sub>COHN</sub> par exemple) les vecteurs ont des orientations différentes et donc échantillonnent presque tout l'espace, quasiment comme dans un spectre de poudre. Les points singuliers de l'histogramme correspondent aux valeurs propres du tenseur d'alignement D<sub>xx</sub>, D<sub>yy</sub> et D<sub>zz</sub>. Pour tracer ce type d'histogramme, il est nécessaire de normaliser tous les couplages dipolaires résiduels mesurés (D<sub>IS</sub>) par rapport aux couplages dipolaires résiduels <sup>1</sup>D<sub>NH</sub> par exemple, car leur gamme de couplage est plus étendue. On obtient le couplage dipolaire résiduel normalisé par rapport NH (D<sub>IS</sub><sup>norm/NH</sup>) à partir du couplage dipolaire résiduel mesuré de la manière suivante :

$$D_{IS}^{norm/NH} = D_{IS} (r_{IS}^3 / \gamma_I \gamma_S) (\gamma_N \gamma_H / r_{NH}^3) \quad (5)$$

Les facteurs d'échelles optimum empiriquement déterminés par Bax et *al* sont présentés dans le tableau suivant.

RDC mesuré	Distance internucléaire effective (Å)	Facteur de normalisation par rapport à NH
<sup>1</sup> D <sub>NH</sub>	1.04	1
<sup>2</sup> D <sub>COHN</sub>	2.05	0.300
<sup>1</sup> D <sub>CON</sub>	1.33	0.120

Tableau : Amplitudes de RDC observés par rapport à  $^1D_{NH}$ 

Cet histogramme permet également de valider les mesures de couplages dipolaires résiduels. En effet si un d'entre eux sort de la distribution, il est préférable de vérifier l'origine de cette valeur extrême (souvent une mauvaise estimation du couplage). Dans ce cas, il est préférable de le sortir des données afin d'éviter d'injecter une erreur expérimentale dans le calcul de structure.

Depuis 1997, l'étude des protéines dans les milieux orientant (bicelles lipidiques, bactériophages filamenteux, mélange de polyéthylène glycols) permet de mesurer les couplages dipolaires résiduels de manière systématique et précise qui donnent des informations orientationnelles à longue portée [120]. Ces derniers ont été utilisés pour différentes applications qui vont de l'affinement de structure à sa détermination *ab initio* [121].

Les protéines paramagnétiques permettent aussi ce type de mesure à l'aide de paramètres tels que les couplages dipolaires résiduels [122], le déplacement chimique paramagnétique [123] ou la corrélation Curie-dipolaire [124].

#### II.3.2.1.2-Les bactériophages filamenteux

Les phages filamenteux ont été introduits par Clore et al et Hansen et al simultanément en 1998 [125, 126]. Ils forment des cylindres de 2  $\mu\text{m}$  de longueur et de 66 Å de diamètre et sont constitués d'un simple brin d'ADN de 7.4 kb entouré d'une gaine de 7620 sous-unités de protéines hélicales identiques de 46 acides aminés chacune. Ces hélices interagissent fortement avec l'ADN et font un angle de 15° avec l'axe du phage. La répétition régulière de ce motif entraîne une forte susceptibilité magnétique à l'origine de l'alignement des phages dans un champ magnétique élevé. Cet alignement se fait parallèlement au champ magnétique  $B_0$ .

Les phages sont stables et posent peu de problème par rapport aux conditions expérimentales de l'échantillon. A pH supérieur à cinq ils sont chargés négativement [ ] et l'alignement induit est majoritairement de nature électrostatique. Des concentrations de sels importantes (~ 100 mM) peuvent être utilisées [127] pour limiter les interactions entre la protéine et les phages. Néanmoins, de trop fortes concentrations en sels entraînent leur

agrégation. Le degré d'alignement obtenu dépend de la concentration de phages et peut être suivi par la mesure du dédoublement quadrupolaire du D<sub>2</sub>O.

## III-MODÉLISATION MOLÉCULAIRE

### III.1-Introduction

La modélisation moléculaire permet une description structurale et dynamique des molécules.

La chimie quantique est la théorie qui explique la liaison chimique et la réactivité chimique. Mais les calculs de mécanique quantique sont applicables seulement sur des systèmes de quelques dizaines d'atomes. La mécanique moléculaire, quant à elle, utilise la mécanique newtonienne pour approcher les structures des molécules. On va modifier les coordonnées des atomes afin d'avoir la conformation de la molécule à l'état d'énergie minimum (en fait, un cliché de la molécule dans des conditions statiques à 0°K). Le modèle représente les atomes dans les molécules par des « boules de mousse électrisées et reliées par des ressorts ». L'ensemble des constantes de force et coefficients d'interactions pour les forces s'exerçant est appelé champ de force dont les paramètres sont basés sur des calculs de spectroscopies, de mécanique quantique et optimisés sur des structures connues. Pour trouver la géométrie optimum d'un ensemble d'atomes, il faut minimiser 3 coordonnées cartésiennes par atome (pour une protéine de 1000 atomes = 3000 coordonnées cartésiennes). Donc, il faut trouver le minimum d'une fonction (l'énergie) dans un espace à quelques milliers de variables. Toutes les méthodes d'optimisation connues ne peuvent trouver qu'un minimum local et on n'est jamais sûr de trouver le minimum global de l'énergie pour une molécule. Il existe des méthodes (Monte-Carlo) qui permettent toutefois d'approcher ce minimum global en franchissant des barrières d'énergie intermédiaires.

La dynamique moléculaire : on sait que les structures ne sont pas figées aux températures auxquelles on veut les étudier. On va donc simuler le mouvement des atomes d'une molécule en intégrant les équations de Newton  $F = ma$  ( $F$  vecteur force,  $a$  : vecteur accélération). Pour un système de milliers d'atomes, toujours modélisés par des « boules de mousse en mouvement électrisées et reliées par des ressorts », on calcule les forces sur

chaque atome, puis leur accélération, et enfin leur vitesse, leurs nouvelles positions, et donc les nouvelles forces. Vu que les forces changent suivant les positions des atomes, il est nécessaire d'effectuer le calcul avec des pas d'intégration petits de l'ordre de la femtoseconde ( $10^{-15}$  seconde). Si on prenait des pas plus grands, la simulation serait plus fautive, car les forces auraient changé pendant le pas d'intégration, et ce changement n'aurait pas été pris en compte.

### III.2-Le calcul de structure

Le calcul de structure devra satisfaire l'ensemble des contraintes expérimentales RMN tout en respectant les contraintes géométriques imposées par la nature des résidus qui composent la protéine. Ainsi la précision et l'exactitude de la structure calculée sont largement déterminées par la qualité de la liste des contraintes RMN. En effet la réduction de l'espace conformationnel, disponible pour la molécule, est due au nombre de contraintes (précision) alors que la consistance des données affectera l'exactitude de la structure. Il est alors nécessaire d'obtenir un maximum de contraintes dites longues portées (faisant intervenir des atomes de résidus éloignés dans la séquence peptidique) qui vont fortement contraindre la structure de la protéine. D'autre part les contraintes nOe, approximatives, ne permettent pas de définir des détails locaux de la structure. Si les contraintes d'angles dièdres, apportent une réponse satisfaisante à ce problème, les couplages dipolaires résiduels se montrent être un bon outil pour palier à ce manque.

Toutes les informations de distances et d'angles dièdres classiquement recueillis constituent une base de données qui servira à définir la position des N atomes de la protéine dans l'espace cartésien au cours du calcul de structure. Toutefois, même si l'ensemble des contraintes expérimentales est très complet, cette base est toujours sous déterminée (ne permet pas de définir de manière non ambiguë les 3N coordonnées des atomes de la protéine). Les informations expérimentales sont alors complétées par un champ de force qui décrit les propriétés géométriques des protéines en terme d'énergie potentielle d'interactions interatomiques liées ou non liées.



### III.2.1-Le champ de force

Le champ de force AMBER, utilisé par le logiciel ARIA<sup>[128, 129, 109]</sup> via le programme CNS<sup>[130]</sup>, est un champ de force semi-empirique, représenté par des termes d'énergie physiques décrivant des interactions atomiques entre atomes liés ou non liés.

#### - Energies potentielles d'interactions liées

Les énergies potentielles liées représentent les énergies correspondant aux déformations de la longueur de liaison, de l'angle entre les liaisons et des angles dièdres de torsion. Leurs expressions sont :

<i>Nature d'interaction</i>	<i>Expression du potentiel</i>
<b>Longueur de liaison covalente</b>	$E_{\text{liaison}} = 1/2 \sum_l k_l (l - l_0)^2$
<b>Angles de liaison</b>	$E_{\text{angle}} = 1/2 \sum_{\theta} k_{\theta} (\theta - \theta_0)^2$
<b>Angles propres de torsion</b>	$E_{\text{propre}} = 1/2 \sum_{\Psi} k_{\Psi} (\Psi - \Psi_0)^2$
<b>Angles impropres de torsion</b>	$E_{\text{impropre}} = 1/2 \sum_{\Phi} k_{\Phi} (\Phi - \Phi_0)^2$

Légende :  $l$  représente la longueur des liaisons covalentes reliant les atomes A-B,  $\theta$  représente l'angle défini par les vecteurs A-B et B-C,  $\Psi$  et  $\Phi$  sont des angles de torsion définis par les plans ABC et BCD.  $l_0$ ,  $\theta_0$ ,  $\Psi_0$  et  $\Phi_0$  sont les valeurs de chacune des grandeurs à l'équilibre. Les  $k_i$  ( $i = l, \theta, \Psi$  ou  $\Phi$ ) sont les constantes de force de chaque force de potentiel.

#### - Energies potentielles d'interactions non liées

Les énergies non liées entre les atomes du système moléculaire peuvent être simulées par deux potentiels différents. Dans un premier temps (au cours du calcul) le potentiel sera défini uniquement à partir de l'interaction de Van der Waals, on parle alors de potentiel

quartique. Puis, dans un second temps (souvent dans la phase de raffinement du calcul) on ajoute la contribution coulombienne qui forme l'interaction électrostatique du potentiel.

$$E_{\text{non-liée}} = E_{\text{van der Waals}} + E_{\text{électrostatique}}$$

#### - La fonction énergie cible

Le principe du calcul de structures de macromolécules sous contraintes est de minimiser une fonction d'énergie cible composée de termes issus du champ de force et d'un terme expérimental.

$$E_{\text{cible}} = E_{\text{liée}} + E_{\text{non-liée}} + E_{\text{exp.}}$$

avec  $E_{\text{exp}} = E_{\text{nOe}} + E_{\text{RDC}}$  où  $E_{\text{nOe}}$  est l'énergie des contraintes nOe et  $E_{\text{RDC}}$  l'énergie des contraintes orientationnelles.

L'énergie expérimentale se décompose en une somme de termes qui correspondent chacun à un type de mesure structurale. Le terme le plus important en RMN est celui des nOe. Les distances sont dérivées du volume des pics nOe à partir de la relation  $D \sim V^{-1/6}$  convertit le volume mesuré en une distance. Pratiquement on introduit des limites de distances basse et haute, L (lower) et U (upper) avec

$$L = \max(0, D - \Delta) \text{ et } U = D + \Delta$$

pour contenir la distance à un intervalle plutôt qu'à une valeur unique. Les contraintes de distances déduites des nOe sont introduites sous forme d'un potentiel harmonique lorsque la distance est proche de la borne inférieure (L) ou supérieure (U) et linéaire lorsqu'on s'en éloigne. La linéarité du potentiel pour de grandes violations permet de ne pas introduire d'énergie trop importante dans le calcul ce qui serait le cas si le potentiel harmonique était conservé. Lorsque la distance interproton est comprise entre [L, D] aucune force n'est appliquée aux atomes.

Les contraintes d'angles dièdres sont également exprimés sous la forme d'un potentiel harmonique.

$$E_{propre} = \sum_{\Psi} \begin{cases} k_{propre}(\Psi - \Psi_{min})^2 & \text{pour } \Psi \leq \Psi_{min} \\ 0 & \text{pour } \Psi_{min} \leq \Psi \leq \Psi_{max} \\ k_{propre}(\Psi - \Psi_{max})^2 & \text{pour } \Psi \geq \Psi_{max} \end{cases}$$

Le traitement des contraintes orientationnelles déterminées à partir des couplages dipolaires résiduels est plus compliqué. En effet, comme nous l'avons déjà vu précédemment, il est nécessaire de définir le tenseur d'alignement de la molécule, qui n'est pas connu *a priori*. Les RDCs sont incorporés au calcul de structure par l'intermédiaire d'un terme d'énergie empirique [tjandra/p96<sup>131</sup>] qui est ajouté aux autres termes conventionnels et qui est de la forme

$$E_{RDC} = k_{rdc}(RDC_{cal} - RDC_{obs})^2$$

$k_{rdc}$  est la force de constante et  $RDC_{cal}$  et  $RDC_{obs}$  sont les valeurs calculées et mesurées des couplages dipolaires résiduels respectivement. Le tenseur d'alignement est considéré comme un repère qui est représenté par une pseudo-molécule localisée dans les environs de la protéine. Cette pseudo-molécule consiste à quatre atomes équidistants (OXYZ). L'un, représente l'origine du repère et les trois autres représentent les trois axes OX, OY, et OZ qui sont orthogonaux les uns par rapport aux autres.  $E_{RDC}$  est évalué en calculant les angles que forme le vecteur considéré avec le tenseur. Durant le processus de calcul, la pseudo-molécule est continuellement réorientée pour permettre la meilleure corrélation entre le couplage dipolaire observé et calculé.

### III.2.2-Méthode de calcul – ARIA1.2

Traditionnellement, la mesure des nOe sert de source primaire pour l'information structurale. ARIA est un logiciel de calcul qui permet d'attribuer les contraintes nOe et de gérer les ambiguïtés d'attribution. Les ambiguïtés sont obtenues au cours de l'attribution

initiale du calcul, à partir d'une liste de contributions possibles pour chaque pic nOe. L'ambiguïté est levée durant le calcul de structure.

#### - ARIA1.2

Aria1.2 effectue un calcul itératif (9 itérations par défaut). Pour chaque itération, on utilise les structures de plus basses énergies du calcul précédent. En fait le protocole d'ARIA repose sur la notion de contrainte structurale ambiguë. Cette approche considère tous les candidats pouvant contribuer à l'intensité du pic de corrélation. Au cours du calcul de dynamique moléculaire, la contribution de chaque candidat est évaluée d'après les distances dans le modèle moléculaire déjà calculé. On peut imaginer deux boules reliées par un ressort et qui sont susceptibles de contribuer à la corrélation. Si leur distance par rapport au modèle est dans la limite  $[L, U]$ , ARIA ne mettra aucune force pour modifier leur distance. Dans le cas où leur distance est hors limite alors l'énergie qui sera donnée à cette contrainte dépendra de la valeur de la distance. Pour une distance trop éloignée ARIA aura tendance à donner une énergie moins grande à cette contrainte (ressort moins fort donc une constante de force plus faible) alors que pour une distance plus proche ARIA donnera plus d'énergie (ressort plus fort donc constante de force plus importante) à cette contrainte qui « tirera » la structure au niveau des deux atomes. Une contrainte ambiguë est satisfaite si au moins un des candidats est compatible selon le modèle.

Comme l'information structurale contenue dans une contrainte ambiguë est plus faible que celle contenue dans une contrainte non-ambiguë, les candidats ayant les plus petites contributions sont progressivement identifiés et éliminés au cours des itérations successives. Toutefois, la combinaison de contraintes ambiguës peut aider à orienter le calcul vers la bonne structure.

ARIA traite indifféremment les pics de corrélation nOe explicites (attribués manuellement) ou implicites (définis seulement par les déplacements chimiques) et chaque itération utilise la même liste de pics de corrélation. Seules les structures (de plus basse énergie) sont héritées de l'itération précédente et affectent les étapes de calibration et d'attributions des pics.

Au début de chaque itération, chaque spectre est calibré indépendamment. Un spectre nOe est rétro-calculé à partir des meilleurs structures de l'itération précédente et comparé au

spectre expérimental afin de déterminer des constantes de calibrations (par types d'atomes). Les volumes expérimentaux sont alors traduits en contraintes de distances. Les meilleures structures de l'itération précédente sont à nouveau analysées pour déterminer les contraintes violées systématiquement. Les bornes de contraintes violées sont ensuite élargies à l'intervalle [0,6] Å et les contraintes toujours violées sont éliminées. Les violations sont définies par deux seuils de tolérance (en distance et en proportion sur l'ensemble structural) qui peuvent être paramétrés par l'utilisateur (paramètres *violtoler* et *violratio*).

L'étape suivante est l'attribution partielle des contraintes ambiguës. Les contraintes acceptées à ce stade sont analysées au regard des meilleures structures de l'itération précédente afin de déterminer la contribution de chaque paire de protons à l'intensité du pic de corrélation. Seules les contributions majoritaires sont conservées. Le paramètre *ambigcutoff* détermine la proportion des contributions conservées.

Après les étapes de calibration et d'attribution, les contraintes des différents spectres NOESY-HSQC sont écrites dans le répertoire de l'itération puis rassemblées (en éliminant les redondances) avant d'être intégrées aux contraintes (angles dièdres, liaisons hydrogènes, RDC) dans un calcul de dynamique moléculaire soit à partir des meilleures structures de l'itération précédente soit à partir de structure aléatoire. Les nouvelles structures serviront de base à l'itération suivante.

## IV-ETUDE DE DYNAMIQUE

D'un point de vue physico-chimique une protéine est un polymère se repliant en une structure tridimensionnelle bien définie en moyenne, non périodique, et maintenue par de interactions faibles non covalentes. Cette structure peut subir de larges fluctuations et se déplacer dans différentes conformations. Les mouvements induits couvrent des échelles de temps très variées (allant de  $10^{-15}$  à  $10^3$  s). Les spins nucléaires sont sensibles aux mouvements des autres noyaux situés à quelques Å ce qui a pour effet d'influencer leur relaxation. L'étude de ce phénomène de relaxation permet d'obtenir des informations de dynamique de la protéine et passe par la mesure RMN de trois paramètres de relaxation : le

vitesse de relaxation longitudinale  $R_1$ , la vitesse de relaxation transversale  $R_2$  et le paramètre de relaxation croisée  $\rho$ .

La dynamique interne des protéines est étudiée essentiellement à partir de la relaxation des noyaux d'azote  $^{15}\text{N}$  ou de carbone  $^{13}\text{C}$ . Pour les deux systèmes de spins  $^{15}\text{N}$  et  $^{13}\text{C}$ , il existe une interaction dipolaire (DD) avec le proton et une anisotropie du déplacement chimique (CSA pour chemical shift anisotropy). Ces deux phénomènes créent un champ magnétique local sur les noyaux qui va provoquer la relaxation des systèmes excités. Ce champ magnétique local varie en fonction de l'orientation de la liaison H-X par rapport au champ magnétique statique. Cette fluctuation est due à la réorientation de la protéine sur elle-même. L'évolution temporelle de l'orientation de la liaison H-X, donc du champ magnétique local, est décrite, mathématiquement, par une fonction de corrélation dont la transformée de Fourier donne la fonction de densité spectrale  $J(\omega)$ , qui permet une description dans le domaine des fréquences du phénomène de relaxation.

La mesure de ces paramètres de relaxation permet d'extraire les paramètres de dynamique tel que le temps de corrélation globale ( $\tau_c$ ), le temps de corrélation des mouvements internes ( $\tau_i$ ), et les paramètres d'ordre  $S^2$ . Le temps de corrélation globale ( $\tau_c$ ) est le temps moyen dont une molécule en solution a besoin pour accomplir une rotation d'un radian (dans la gamme des ns). Le temps de corrélation local ( $\tau_i$ ) (de la ps à la ns) a la même signification que  $\tau_c$  mais concerne un vecteur (liaison H<sup>N</sup>-N par exemple). Les paramètres d'ordre ( $S^2$ ) donnent une mesure de l'amplitude des mouvements internes des vecteurs étudiés. Plus sa valeur tend vers 1, plus le mouvement de la liaison est restreint.

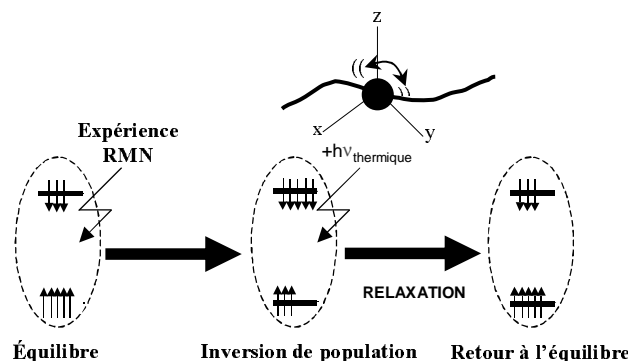


Figure 16 : Schéma très simplifié du phénomène de relaxation

## Chapitre III

### Production et attribution de la protéine SufA

La production d'une protéine soluble et stable dans le temps est une étape primordiale. Cependant, cette étape est le plus souvent consommatrice de temps et d'énergie. Les progrès des outils de biologie moléculaire, ainsi que le développement incessant des techniques de purification, ont grandement facilité la réalisation de tel projet. L'obtention d'un échantillon protéique dans les conditions requises pour une étude par RMN, est suivie de l'étape d'attribution. Ces deux étapes constituent la porte d'entrée de l'étude structurale par RMN qui nécessite ensuite l'extraction de contraintes structurales RMN qui sont introduites dans le calcul de structure. Notons que pour pouvoir extraire les données structurales et dynamiques par RMN, il faut connaître les déplacements chimiques de tous les atomes de la protéine. En cela, l'étape d'attribution est fondamentale pour la réussite du projet.

#### I-PRÉPARATION DE L'ÉCHANTILLON DE LA PROTÉINE SufA

La protéine SufA, possédant une étiquette Histidine (tag) en position C-terminale, a été caractérisée biochimiquement et spectroscopiquement dans le cadre d'étude au laboratoire. Le protocole de production de la protéine SufA, présenté dans le chapitre « matériels et méthodes » de cette thèse, était donc déjà établi pour ces recherches. La protéine, identifiée comme étant un dimère, imposait de réaliser un double marquage  $^{15}\text{N}$ ,  $^{13}\text{C}$  pour les raisons que nous avons indiquées dans le chapitre précédent. Toutefois, afin de contrôler la faisabilité de l'étude par RMN, nous avons tout d'abord effectué un simple marquage  $^{15}\text{N}$  uniforme de la protéine SufA. D'autre part, le protocole de production de la protéine marquée a été adapté pour le marquage, et est identique pour les deux types de marquages.

### I.1-Résultats de production de la protéine SufA

1) L'analyse par SDS-PAGE des cellules surexprimant SufA a montré une expression importante d'un polypeptide d'environ 14 kDa, absent dans les cellules non induites.

2) La pureté de la protéine après colonne NiNTA puis colonne Superdex75, analysée par SDS-PAGE (Figure 2), montre que SufA est pure.

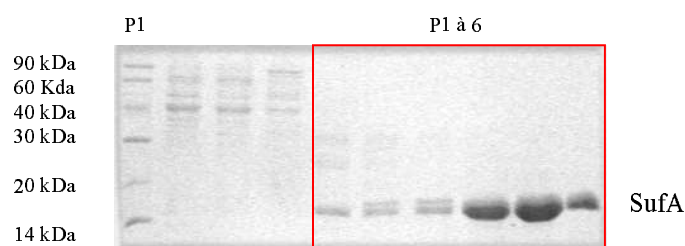


Figure 2: Gel d'électrophorèse 15 % - analyse de la purification après Sdx-75  
Puits 1 : marqueurs de poids moléculaires - Puits 1 à 6 : fractions de purification

3) Bilan de la purification de la SufA d'*E. coli* à partir de 3 litres de culture :

	Quantité (mg) de protéines	Rendement <sup>a</sup> (%)
Superdex 75	22	15

(a) Rendement =  $[Q^{\text{té}}(\text{mg}) \text{ protéines avant purification} / Q^{\text{té}}(\text{mg}) \text{ protéines après purification}] \times 100$

## II-CARACTÉRISATION PAR RMN DE SufA

Il est intéressant, avant de démarrer une analyse RMN, d'effectuer une analyse « bioinformatique » de la protéine d'intérêt. En ayant recours à des logiciels de prédiction de



structures secondaires, et/ou en effectuant des alignements de séquences afin de trouver des protéines homologues dont la structure est connue. On peut ainsi obtenir des informations préliminaires quant à la structuration de la protéine. En ce qui concerne la protéine SufA, nous avons connaissance de la structure de la protéine homologue IscA sous la forme apo, obtenue par radiocristallographie [<sup>132</sup>].

## II.1-Caractéristiques de SufA d'*E. coli*

La caractérisation de la protéine SufA a été réalisée en utilisant le serveur de biologie moléculaire « Expasy ». On extrait un certain nombre de caractéristiques physico-chimiques de la protéine par l'utilisation d'outils d'analyse (« ProtParam ») de la séquence primaire disponibles sur le site. internet : <http://www.expasy.org/cgi-bin/protparam>.

*Séquence :*

**MDMHSGETFNPQDFAWQGLTLPAAAIHIRELVAKQPGMVGVR LGVKQTGCAGFGYVLDSV  
SEPKDDLLFEHDGAKLFVPLQAMPFIDGTEVDFVREGLNQIFKFHNPKAQNECGCGESF  
GVLEHHHHHH**

*Nombre d'acides aminés : 122*

*Masse moléculaire : 13300.1 Da (M<sub>Expasy</sub>)*

*pI théorique (point isoélectrique) : 4.85*

*Composition en acides aminés :*

<b>Ala (A)</b>	<b>9</b>	<b>7.4%</b>
<b>Arg (R)</b>	<b>3</b>	<b>2.5%</b>
<b>Asn (N)</b>	<b>4</b>	<b>3.3%</b>
<b>Asp (D)</b>	<b>9</b>	<b>7.4%</b>
<b>Cys (C)</b>	<b>3</b>	<b>2.5%</b>
<b>Gln (Q)</b>	<b>7</b>	<b>5.7%</b>
<b>Glu (E)</b>	<b>7</b>	<b>5.7%</b>
<b>Gly (G)</b>	<b>14</b>	<b>11.5%</b>
<b>His (H)</b>	<b>4</b>	<b>3.3%</b>
<b>Ile (I)</b>	<b>4</b>	<b>3.3%</b>
<b>Leu (L)</b>	<b>10</b>	<b>8.2%</b>
<b>Lys (K)</b>	<b>6</b>	<b>4.9%</b>
<b>Met (M)</b>	<b>4</b>	<b>3.3%</b>
<b>Phe (F)</b>	<b>10</b>	<b>8.2%</b>
<b>Pro (P)</b>	<b>7</b>	<b>5.7%</b>
<b>Ser (S)</b>	<b>4</b>	<b>3.3%</b>
<b>Thr (T)</b>	<b>5</b>	<b>4.1%</b>
<b>Trp (W)</b>	<b>1</b>	<b>0.8%</b>
<b>Tyr (Y)</b>	<b>1</b>	<b>0.8%</b>
<b>Val (V)</b>	<b>10</b>	<b>8.2%</b>

De cette analyse on en déduit que : la protéine SufA est constituée de 31.2% de chaînes latérales contenant des groupes méthyles qui sont susceptibles de former une ou plusieurs zones hydrophobes et de 13.1% de résidus aromatiques.

**II.2-Prédiction de structures secondaires de SufA**

La prédiction de structure secondaire a été réalisée à l'aide du programme PHDsec ([http://cubic.bioc.columbia.edu/predictprotein/submit\\_def.html](http://cubic.bioc.columbia.edu/predictprotein/submit_def.html)).

D'après la prédiction de structure secondaires, plus de 40% des résidus de la protéine sont impliqués dans un élément de structure secondaire, avec environ 25% de brins  $\beta$  et 15% d'hélices  $\alpha$ .

On constate sur la Figure 1 ci-dessous que les structures secondaires issues de la prédiction de structure de SufA sont identiques à celles d'IscA de structure connue. Comme les deux protéines présentent une forte identité de séquence, on peut supposer que la structure de SufA sera très proche de celle d'IscA.

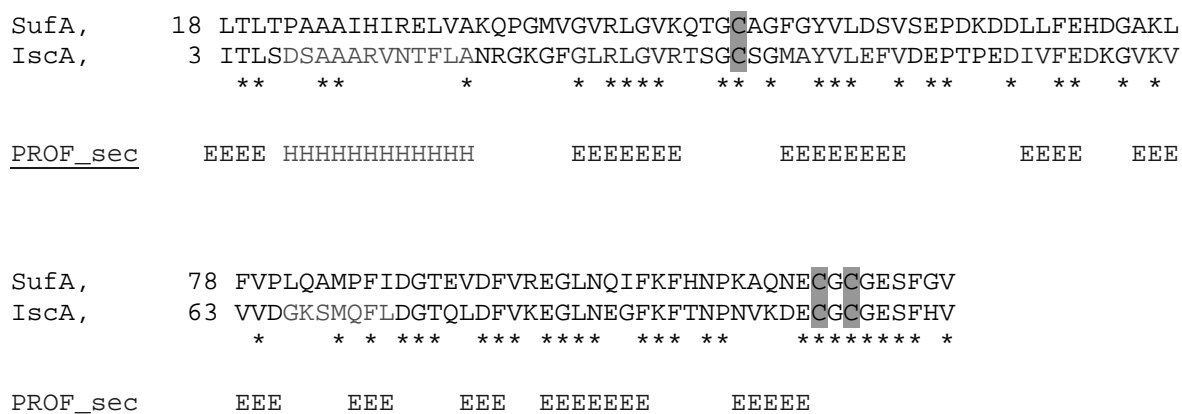


Figure 1 : Alignement de séquences et prédiction de structures secondaires des protéine SufA et IscA d'*E. Coli*

(48.6% identité pour 105 résidus)

E = feuillet  $\beta$  et H = hélice  $\alpha$

### II.3-ÉCHANTILLON RMN

Les échantillons préparés pour l'étude de SufA ont une concentration de l'ordre du millimolaire (mM) et sont stables dans le temps (plusieurs semaines à 4°C). Ces caractéristiques se prêtent donc bien à l'étude RMN. L'échantillon est dans un tampon phosphate 50 mM de pH égal à 6.8, en présence de dithioréitol (DTT) de concentration 10 mM afin d'éviter la présence d'oligomères par formation de ponts disulfures. On ajoute, d'autre part, un cocktail antiprotéases de type COMPLETE<sup>TM</sup> (Boehringer) pour éviter la protéolyse de la protéine, ainsi que de 10 % (du volume final) de D<sub>2</sub>O nécessaire au verrouillage du signal RMN. On injecte ensuite un gaz inerte, de l'argon, dans le tube RMN où se trouve l'échantillon de protéine, ainsi préparée, de 450  $\mu$ L. Le tube RMN est alors scellé afin de protéger l'échantillon (oxygène, contaminants, ...).

Échantillon	<sup>15</sup> N	<sup>15</sup> N/ <sup>13</sup> C
Concentration	1 mM	1 mM
Volume	450	450
D2O stérile	10%	10%
Tampon	50 mM phosphate, pH = 6.8	50 mM phosphate, pH = 6.8
Antiprotéases	Oui	Oui
DTT	10 mM	10 mM

Notons que toutes les expériences ont été enregistrées à 25°C

### III-ATTRIBUTION DES RÉSONANCES

Un spectre  $^1\text{H}$ - $^{15}\text{N}$  HSQC constitue l'empreinte de la protéine et permet d'estimer la faisabilité d'une étude par RMN de la protéine. Dans ce spectre, pour chaque acide aminé on trouve une corrélation  $^1\text{H}$ - $^{15}\text{N}$  et la dispersion des pics de corrélation est représentative de l'état de repliement de la protéine. Une protéine bien repliée et structurée est caractérisée par des fréquences bien étalées dans la gamme de résonance des NH (de 6 à 12 ppm en fréquence protons). La distribution de l'intensité des pics est représentative de la flexibilité relative de la protéine. Ainsi les pics intenses peuvent provenir de parties flexibles de la protéine ou de petits contaminants. Le spectre HSQC permet également de vérifier très rapidement l'homogénéité de la préparation par le décompte des pics de corrélation. La présence de formes mineures chimiques ou conformationnelles peut être ainsi détectée. Le spectre  $^1\text{H}$ - $^{15}\text{N}$  HSQC de la protéine SufA est présenté sur la Figure 3.

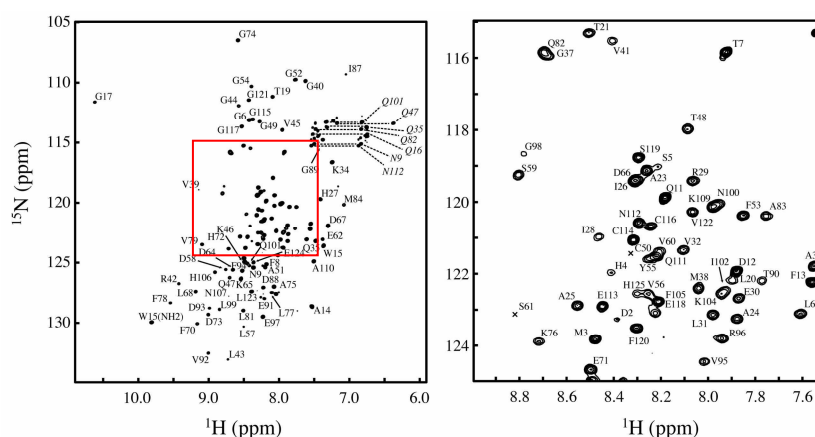


Figure 3 : Spectre  $^1\text{H}$ - $^{15}\text{N}$  HSQC de SufA doublement marquée  $^{13}\text{C}$  et  $^{15}\text{N}$ .

On remarque que ce spectre  $^1\text{H}$ - $^{15}\text{N}$  HSQC de la protéine SufA est très bien étalé en fréquence, avec des intensités de pics relativement homogène, suggérant une protéine bien structurée. On distingue nettement les NH et NH<sub>2</sub> (entre 6.5 et 7.5 ppm en proton, entre 110 et 115 ppm en azote) des résidus de types Arg, Asn et Gln, ainsi que des pics de corrélation qui sortent aux alentours de 10.5 à 11 ppm. D'autre part il n'y a pas de dédoublement des pics, comme aucun autre pic parasite, ce qui confirme que l'on est en présence d'un échantillon pure et homogène.

Suite à ce résultat, il a été décidé de poursuivre l'étude par RMN de la protéine SufA par l'enregistrement d'expériences hétéronucléaires, nécessaires à l'attribution de la chaîne polypeptidique, c'est-à-dire un étiquetage par le déplacement chimique de l'ensemble des atomes qui constituent la protéine.

### III.1-Stratégie d'attribution des résonances de la chaîne principale

La méthode d'attribution des fréquences de résonance des atomes de la chaîne principale consiste en une série d'expériences de triple résonance qui permet de relier les fréquences des noyaux  $H_i$  et  $N_i$  du résidu ( $i$ ) aux fréquences des carbones du résidu précédent ( $i-1$ ) et aux fréquences des carbones du même résidu  $i$ .

Les expériences 3D triples résonances utilisées pour l'attribution de SufA sont : HNCA, CACBNH, CACB(CO)NH, HNCO, HN(CA)CO.

	Informations obtenues	Sensibilité de l'expérience	Durée d'expérience
<b>HSQC <math>^1H</math>-<math>^{15}N</math></b>	$H_{Ni}$ , $N_i$	++++	
<b>HNCA</b>	$[H_{Ni}, N_i, C_{\alpha i}]$ et $[H_{Ni}, N_i, C_{\alpha i-1}]$	++	72 Heures
<b>CBCANH</b>	$[H_{Ni}, N_i, C_{\alpha i}, C_{\beta i}]$ et $[H_{Ni}, N_i, C_{\alpha i-1}, C_{\beta i-1}]$	+	72 Heures
<b>CB(CA)CO NH</b>	$H_{Ni}$ , $N_i$ , $C_{\alpha i-1}$ , $C_{\beta i-1}$	++	72 Heures
<b>HNCO</b>	$H_{Ni}$ , $N_i$ , $CO_{i-1}$	+++	72 Heures
<b>HN(CA)CO</b>	$[H_{Ni}, N_i, CO_i]$ et $[H_{Ni}, N_i, CO_{i-1}]$	+	72 Heures

La première expérience, faisant suite à l'HSQC $^1H$ - $^{15}N$ , donnent accès aux fréquences de résonance des  $C_{\alpha i}$  et  $C_{\alpha i-1}$ . Les deux suivantes permettent de faire des connexions entre les  $C_{\beta i}$ ,  $C_{\beta i-1}$ ,  $C_{\alpha i}$ ,  $C_{\alpha i-1}$  alors que les deux dernières permettent d'observer les  $CO_i$  et  $CO_{i-1}$ . Les trois premières expériences donnent donc une information redondante du fait que l'on édite plusieurs fois le déplacement chimique des  $C_{\beta i}$ ,  $C_{\beta i-1}$ ,  $C_{\alpha i}$ ,  $C_{\alpha i-1}$  d'un résidu  $i$ . Il en est de même pour les deux dernières mais pour les  $CO_i$  et  $CO_{i-1}$ . Cette redondance d'information permet d'effectuer une attribution séquentielle de la chaîne principale comme sur l'exemple de la Figure 4. Cette attribution peut s'effectuer manuellement mais il existe aujourd'hui des

logiciels permettant d'effectuer cette tâche de manière semi-automatique, tel que ALPS que nous avons utilisé pour l'analyse de la protéine SufA.

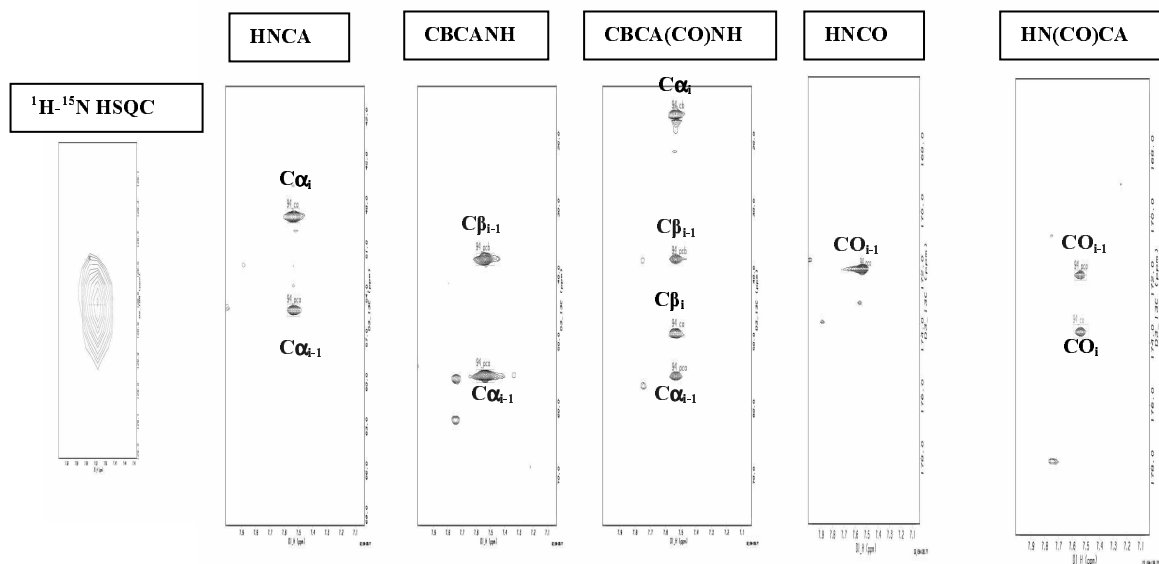


Figure 4 : Exemple de coupes pour l'attribution séquentielle.

Partant d'un pic de corrélation ( $^1\text{H}$ - $^{15}\text{N}$ )<sub>i</sub> les expériences 3D dont on présente des coupes dans la dimension  $^{15}\text{N}$  permettent d'identifier et de connecter les résidus. On trouve de l'information redondante entre les différentes expériences, ceci permet de connecter un résidu *i* avec le résidu *i*-1 (précédant). Par exemple l'expérience HN(CO)CA permet d'identifier le CO<sub>i</sub> et le CO<sub>i-1</sub>. La discrimination entre les deux pics se fait à partir de l'expérience HNCO qui ne donne que le CO<sub>i-1</sub>. En appliquant la même procédure aux autres expériences, on identifie Cα<sub>i</sub>, Cα<sub>i-1</sub>, Cβ<sub>i</sub>, et Cβ<sub>i-1</sub>.

### III.2-Attribution séquentielle et CSI :

L'attribution séquentielle des fréquences consiste comme on vient de le voir, en la reconstitution d'un puzzle. Dans le cas idéal, pour chaque couple H<sub>Ni</sub>, N<sub>i</sub> nous connaissons le CO<sub>i</sub>, CO<sub>i-1</sub>, Cβ<sub>i</sub>, Cβ<sub>i-1</sub>, Cα<sub>i</sub>, Cα<sub>i-1</sub> alors si nous trouvons un résidu H<sub>Nj</sub>, N<sub>j</sub> avec CO<sub>j</sub> = CO<sub>i-1</sub>, Cα<sub>j</sub> = Cα<sub>i-1</sub>, Cβ<sub>j</sub> = Cβ<sub>i-1</sub>, on peut dire que le résidu *j* précède le résidu *i*. De cette manière, nous pouvons reconstituer les connectivités séquentielles au fur et à mesure. L'algorithme du programme ALPS (Assignment for Labelled Protein Spectra), développé au laboratoire RMN de l'IBS [98, 133], se base sur la liste des fréquences de résonances déterminées précédemment à partir des expériences de triple résonance. Il construit des pseudo-résidus et utilise une recherche systématique pour soumettre une attribution et une optimisation par recuit simulé. La recherche systématique se base sur la séquence protéique et sur des plages de déplacement chimique des noyaux C<sub>α</sub> et C<sub>β</sub> pour les différents acides aminés. ALPS propose en sortie une

ou plusieurs connectivités pour le résidu  $i$  et le type d'acide aminé qui peut correspondre aux résidus  $i$  et  $i-1$ . L'attribution semi-automatique effectuée par ALPS nécessite toutefois une intervention de l'utilisateur pour l'aider à lever les ambiguïtés restantes. Le logiciel effectuera alors une évaluation de la proposition qui lui est faite.

L'application de cette analyse semi automatique sur les fichiers de pics issus de l'analyse avec le logiciel FELIX de tous les spectres hétéronucléaires de SufA a permis l'attribution de 93.5% des résidus de la protéine SufA (130 – 7 prolines = 123 résidus attribuables dont Q16, L18, F86, F103 non attribués ainsi que 4 His de l'étiquette Tag).

Cette attribution qui ne concerne que les atomes  $H^N$ ,  $CO$ ,  $C_\alpha$  et  $C_\beta$  de la chaîne polypeptidique est une étape importante dans le cadre de l'analyse structurale par RMN. En effet à partir de ces données de déplacements chimiques nous pouvons déjà obtenir de l'information structurale à propos de la protéine. Pour cela, Wishart et *al* ont développé une méthode d'analyse des déplacements chimiques nommée index de déplacements chimiques (CSI pour chemical shift index) [134]. C'est une méthode statistique qui se base sur une analyse des valeurs expérimentales des déplacements chimiques  $^{13}C$ , ceux-ci étant très sensibles à leur environnement tant chimique que structural, donc aux éléments de structures secondaires. Les déplacements chimiques expérimentaux et l'index de déplacement chimique des  $^{13}C_\alpha$ ,  $^{13}CO$  et  $^{13}C_\beta$  sont comparés puis un consensus est effectué. La valeur des variations de déplacements chimique  $[\delta(C_\alpha + C_\beta - CO)_{exp} - (C_\alpha + C_\beta - CO)]$  est analysée comme suit : à une valeur positive élevée on affecte le chiffre +1, à une valeur négative élevée on affecte le chiffre -1. Le long de la séquence protéique, une suite d'au moins quatre +1 sans alternance avec un -1 suggère la présence d'une hélice  $\alpha$ , et inversement une suite d'au moins quatre -1 sans alternance avec un +1 suggère la présence d'un brin  $\beta$ .

L'analyse CSI de SufA sur la base des déplacements chimiques du  $CO$ ,  $C_\alpha$  et du  $C_\beta$  est présentée à la Figure 5.

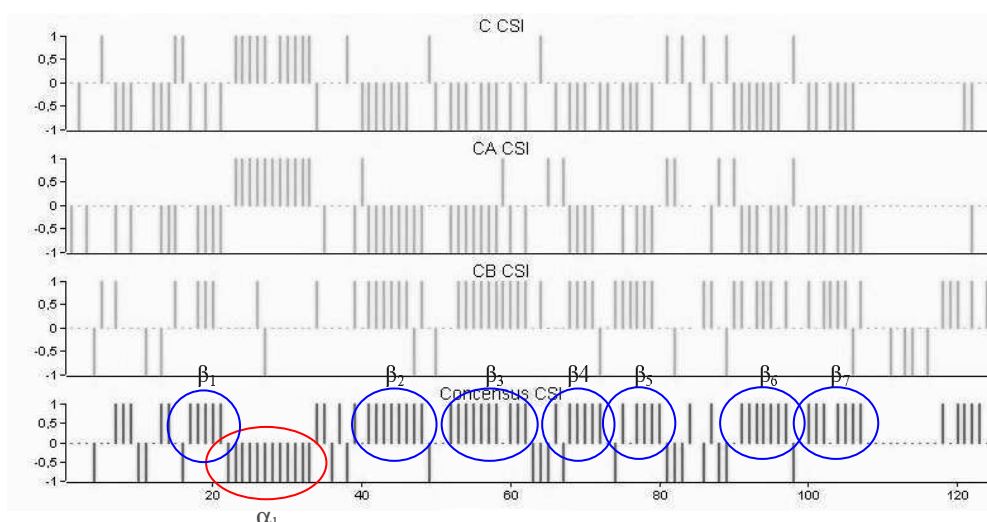


Figure 5 : Analyse des éléments de structures secondaires de SufA à partir du CSI et des déplacements chimiques de la protéine. La protéine SufA possède 7 brins  $\beta$  et une hélice  $\alpha$ . Il n'y a, par contre, aucun consensus en ce qui concerne l'éventuelle petite hélice  $\alpha$  qu'on observe sur la protéine homologue IscA entre les brins  $\beta_5$  et  $\beta_6$ .

On constate que la protéine SufA possède 7 brins  $\beta$  et une hélice  $\alpha$  du côté N-terminal. Ce résultat corrèle parfaitement avec la prédiction de structure secondaire (page 71) ce qui nous autorise à dire que les éléments de structures secondaires sont maintenant connus. Cette structure est pratiquement identique à la structure de la protéine homologue IscA mis à part la petite hélice  $\alpha_2$  (qui se situe au niveau des résidus 66-72 d'IscA). Ceci ne veut pas dire qu'elle n'existe pas chez SufA. En effet l'analyse CSI nécessite que quatre résidus au moins aient le même profil statistique. Or la petite hélice  $\alpha_2$  n'est constituée que de trois résidus et l'un d'entre eux, la phénylalanine 86, n'est pas attribué.

### III.3-Analyse de mutants :

Les mutants de la protéine SufA analysés par RMN correspondent aux mutations (C→S) de la protéine au niveau des trois cystéines conservées 50, 114 et 116. L'objectif de cette étude a été d'analyser les éventuelles modifications structurales au sein des différents mutants. On espérait ainsi observer l'impact de ces mutations sur la structure de la protéine SufA et finalement déterminer si il y avait lors d'une mutation des effets sur les deux autres cystéines non mutées.

Pour cela, trois spectres  $^1\text{H}$ - $^{15}\text{N}$  HSQC ont été enregistrés à 600 MHz. La procédure de marquage et de purification des mutants fut totalement identique à celle de la protéine

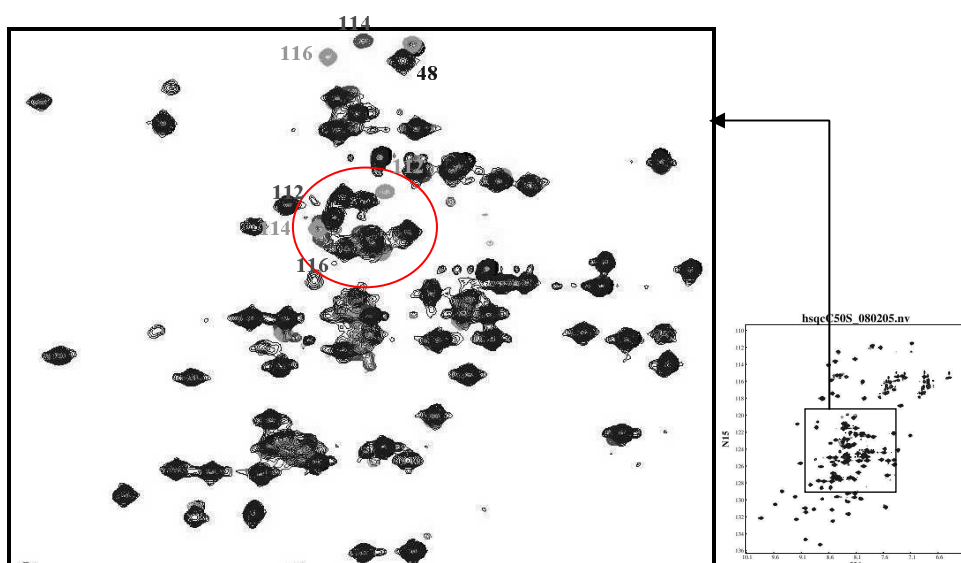


sauvage ainsi que la préparation de l'échantillon RMN. Les trois échantillons avaient une concentration d'environ 1mM.

Les spectres HSQC ont été enregistrés avec les paramètres d'acquisition et de traitement identiques à ceux de la protéine sauvage.

*-série de spectres 1H-15N HSQC*

a)



b)



Figure 6 :

a) Superposition des spectres 1H-15N HSQC (zoom entre 7.5 ppm/9 ppm dans la dimension  $^1\text{H}$  et 119 ppm/127 ppm dans la dimension  $^{15}\text{N}$ ) des trois mutants cystéines avec la protéine sauvage.

En bleu C50S, en rouge C114S, en vert C116S et en noir la protéine SufA sauvage.

b) Superposition des spectres 1H-15N HSQC (zoom entre 7 ppm/9 ppm dans la dimension  $^1\text{H}$  et 110 ppm/119 ppm dans la dimension  $^{15}\text{N}$ ) des trois mutants cystéines avec la protéine sauvage.

En bleu C50S, en rouge C114S, en vert C116S et en noir la protéine SufA sauvage.

On remarque que :

- La mutation C114S affecte les résidus 112, 116, 115 et 117.
- La mutation C116S affecte les résidus 112, 114, 115 et 117.
- La mutation C50S affecte les résidus 48, 49 et 52.

On conclue de ces expériences que : l'intégrité de la protéine n'a pas été affectée par les trois différentes mutations. En effet, on n'observe (figures 6a et 6b) aucune modification significative des déplacements chimiques sur l'ensemble des résidus, mis à part ceux qui entourent les cystéines 114 et 116 et les résidus autour de la cystéine 50. Les mutations des cystéines 114 et 116 n'affectent pas la structure globale de la protéine dans l'environnement de la cystéine 50 et inversement.

Cela semble plus évident si on superpose par exemple le spectre  $^1\text{H}$ - $^{15}\text{N}$  HSQC du mutant C116S avec la protéine SufA sauvage pour les mêmes zones de spectre (Figure 7)

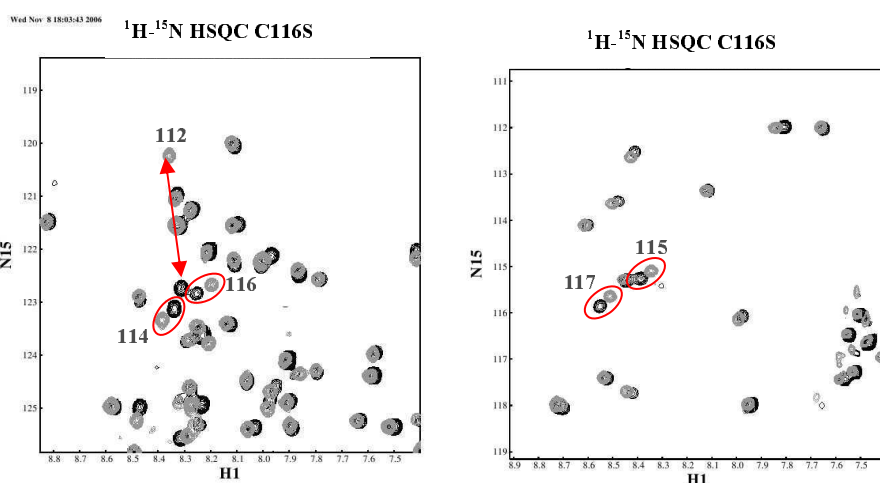


Figure 7 : Superposition du spectre  $^1\text{H}$ - $^{15}\text{N}$  HSQC du mutant 116 avec celui de la protéine sauvage dans les mêmes domaines spectraux que ceux présentés ci-dessus.

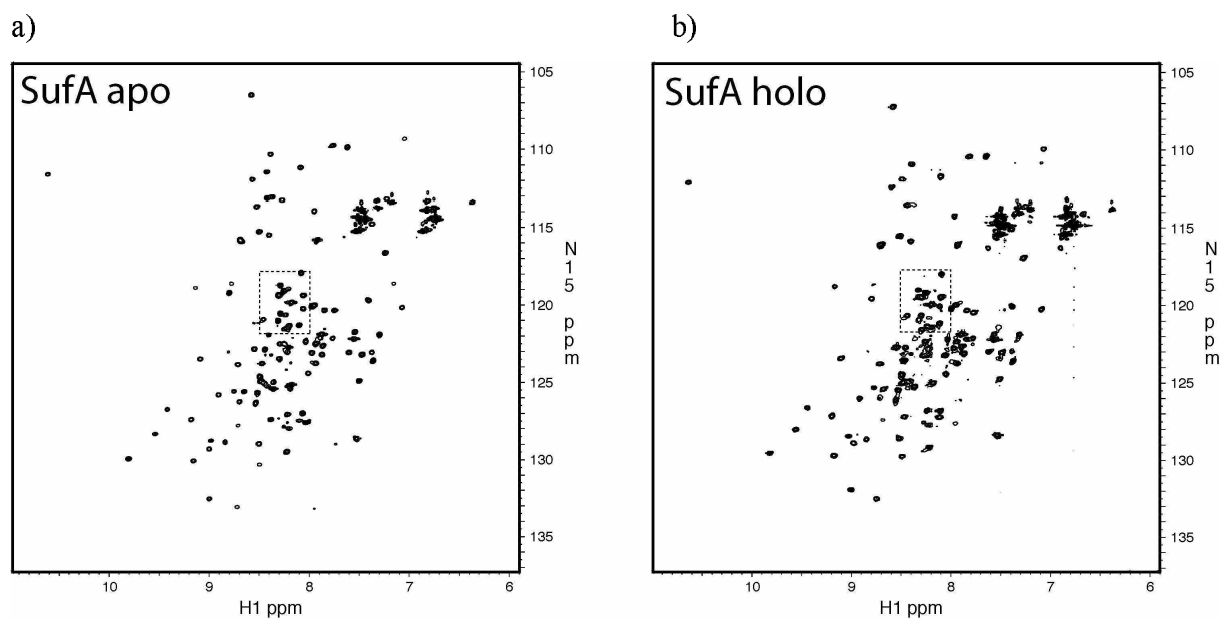
#### III.4-Analyse de la protéine SufA avec son centre [Fe-S]

Un autre type d'expérience réalisée sur la base de l'attribution de la chaîne polypeptidique est l'enregistrement de la protéine SufA avec son centre [Fe-S]. L'objectif est toujours de pouvoir observer une modification structurale de la protéine en analysant les déplacements chimiques des résidus attribués de la protéine SufA.

La protéine SufA est produite et purifiée selon le même protocole que décrit dans le paragraphe « matériels et méthodes ». La différence vient du fait que l'on introduit chimiquement le centre [Fe-S] dans la protéine étant donné que celle-ci est purifiée sous forme apo. La reconstitution du centre [Fe-S] a été effectuée en anaérobiose en boîte à gant selon un protocole décrit dans la partie matériels et méthodes.

Les spectres  $^1\text{H}$ - $^{15}\text{N}$  HSQC sont enregistrés avec des paramètres d'acquisition et de traitement identiques à ceux des expériences précédentes. Il est important de noter que l'échantillon préparé a été préalablement analysé par des méthodes biochimiques et spectroscopiques (dosage du fer, UVvis) pour vérifier que le centre [Fe-S] était intègre.

*-Spectre  $^1\text{H}$ - $^{15}\text{N}$  HSQC de la protéine SufA holo*



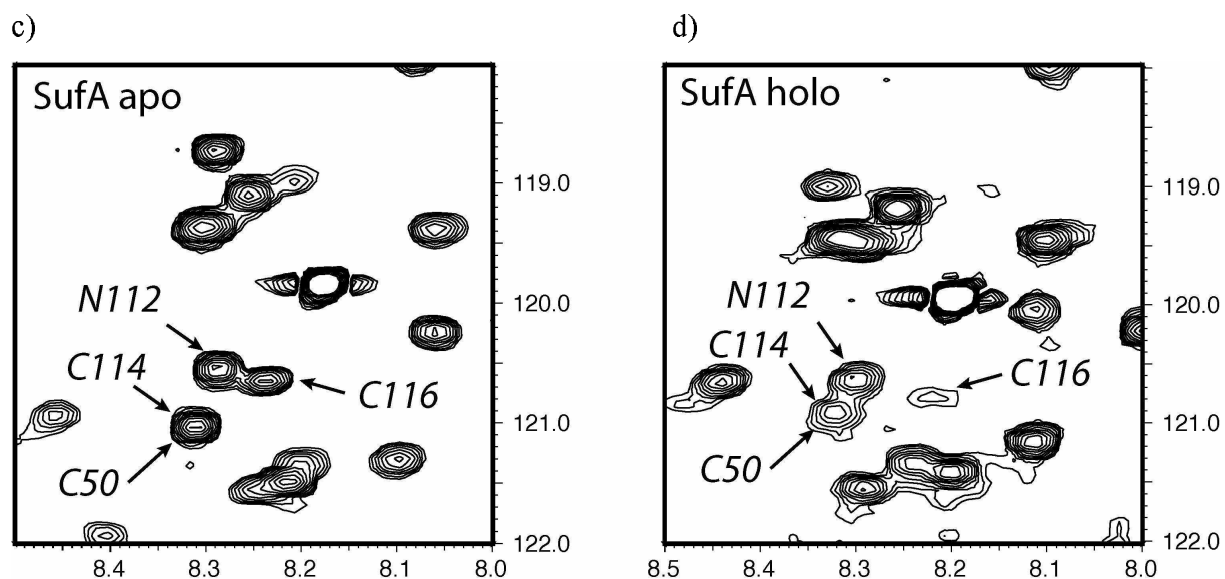


Figure 8:

a) et b) Comparaison de deux spectres HSQC de la protéine SufA apo et holo. Enregistrés à 600 MHz, 25°C. On constate que la protéine conserve globalement le même repliement avec et sans centre [Fe-S].

c) et d) Zoom de la région des cystéines conservées 50, 114 et 116. La présence du centre [Fe-S] affecte les corrélations  $^1\text{H}$ - $^{15}\text{N}$  des cystéines pour lesquelles on observe une diminution d'intensité du pic.

Comme pour les mutants, on observe que l'intégrité de la protéine SufA est conservée. Par contre, d'une manière générale on voit que les pics de résonance sont plus larges (figure 8a) dans la forme holo. Ceci est dû au fait que le centre métallique de par ses propriétés paramagnétiques affecte la relaxation et le déplacement chimique d'une partie des protons de la protéine. Ce paramagnétisme se fait ressentir dans une sphère d'environ 6 Å et, en fonction du type de contact que le proton a avec le centre métallique : soit seule sa largeur spectrale est affectée (pseudo-contact) soit elle est affectée mais avec en plus un pic fortement déplacé (jusqu'à 100 ppm, contact). En effet, la présence d'un ou plusieurs électrons non appariés dans un composé paramagnétique donne une contribution supplémentaire au déplacement chimique d'un noyau observé en RMN. Cette contribution paramagnétique est de deux natures : le déplacement dipolaire (ou de pseudo contact)  $\delta_{\text{dip}}$  à travers l'espace provenant du champ dipolaire créé par ce(s) électron(s) au niveau du noyau observé et le déplacement de contact (ou de Fermi)  $\delta_{\text{con}}$  reflétant la délocalisation à travers les liaisons chimiques de la densité de spin électronique sur les noyaux. Le paramagnétisme, du au contact, n'affecte donc que les résidus qui ont une liaison covalente avec le centre métallique. Quand on analyse le spectre  $^1\text{H}$ - $^{15}\text{N}$  HSQC de la protéine SufA holo (Figure 8d) on remarque que l'intensité de la cystéine 116 est beaucoup plus faible. Les cystéines 50 et 114 ont des corrélations superposées mais on

voit bien que celles-ci sont aussi plus faibles. On peut donc raisonnablement émettre l'hypothèse que ces trois cystéines sont très proches du site actif. La Figure 9 représente la différence des déplacements chimiques des  $^1\text{H}$  amidiques entre les formes apo et holo de la protéine SufA. Cet histogramme montre d'une part que globalement la présence du centre métallique affecte peu les systèmes de spin en terme de déplacement chimique (la plus grande valeur déterminée est  $< 0.1$  alors qu'on considère que cela devient significatif pour une valeur  $> 0.3$ ), mis à part les cystéines conservées plus un certain nombre de résidus autour de celles-ci. Il semble évident que la boucle où se situe la cystéine 50, comme la partie C-terminale où on trouve les cystéines 114 et 116 sont proches du site actif.

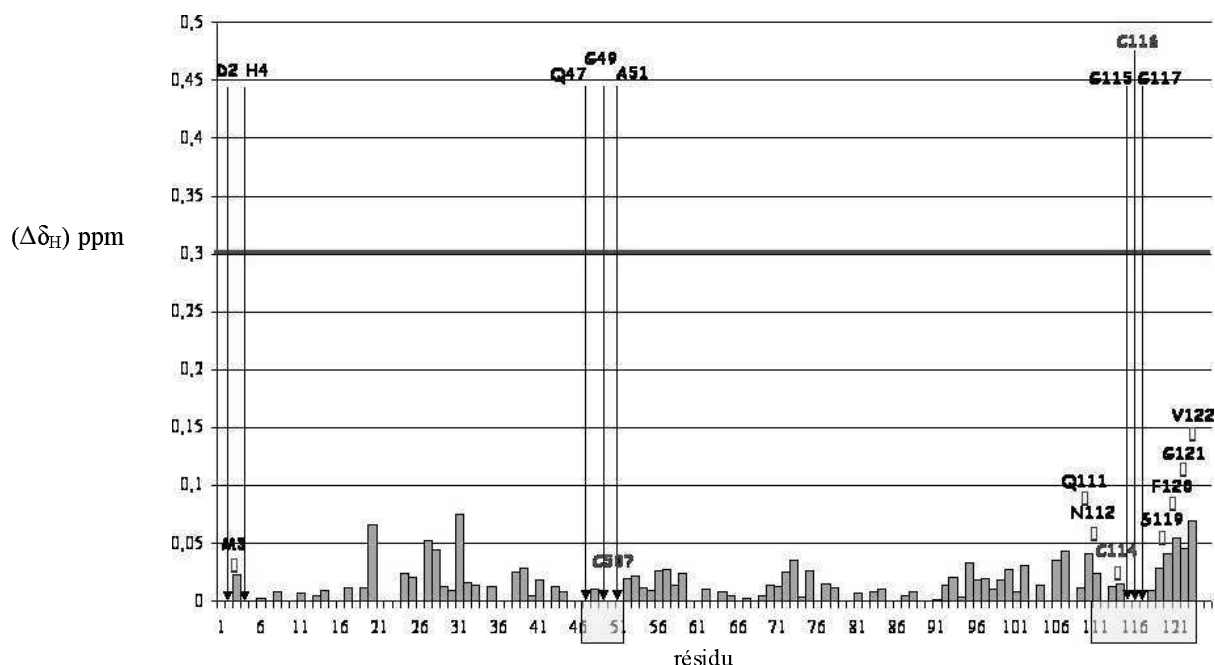


Figure 9 : Différences de déplacements chimiques (barres violettes) des protons  $\text{H}^{\text{N}}$  de la chaîne principale entre la forme apo et la forme holo de la protéine SufA. En ordonnée on donne  $(\Delta\delta_{\text{H}})$  en ppm.

Sur cet histogramme de variation des déplacements chimiques, on fait apparaître (rectangle jaune) les résidus affectés par le paramagnétisme de centre  $[\text{Fe-S}]$ .

On observe de cette expérience que :

- La variation de déplacement chimique est faible pour l'ensemble des résidus. Ceci tend à montrer que le repliement global de la protéine SufA est conservé.

- En ce qui concerne les résidus notés, l'intensité de leur signal a fortement diminuée. Les résidus 111-122 disparaissent pratiquement du spectre ainsi que les résidus 47-51. Les intensités résiduelles que l'on observe sur le spectre de la protéine SufA holo sont

certainement dues à de l'apo protéine. Les pics des résidus affectés par le paramagnétisme du centre [Fe-S] sont trop larges ou complètement déplacés, hors de la gamme de déplacement chimique, pour pouvoir être observés.

Les régions des trois cystéines semblent donc bien affectées par le paramagnétisme du centre [Fe-S]. Toutefois, si on peut conclure que les cystéines sont proches du site actif, on ne peut pas affirmer avec ces résultats qu'elles sont liées au Fer directement de manière covalente.

### III.5-Attribution des chaînes latérales

Différentes stratégies sont habituellement utilisées pour l'attribution des fréquences proton et carbone des chaînes aliphatiques [<sup>135</sup>]. La première est basée sur l'expérience HCCH-TOCSY qui corrèle entre eux les carbones et les protons de la chaîne latérale. Dans ce cas, l'attribution complète d'un résidu particulier se fait par identification des déplacements chimiques des carbones C<sub>α</sub> et C<sub>β</sub>, obtenus lors de l'attribution de la chaîne principale. Cette stratégie présente l'inconvénient de s'appuyer sur les fréquences des carbones et protons aliphatiques issues des expériences CBCA(CO)NH ou HBHA(CO)NH, qui ont une probabilité de recouvrement élevée. La seconde stratégie consiste à corrélérer directement les déplacements chimiques des protons et carbones aliphatiques au déplacement chimique de l'amide correspondant grâce aux expériences (H)CC(CO)NH-TOCSY et H(CC)CONH-TOCSY. Par contre, l'utilisation de ces deux expériences ne permet pas de déterminer quel proton est lié à quel carbone.

Nous avons utilisé une combinaison de ces deux stratégies pour attribuer les chaînes latérales de SufA. Elle repose sur l'acquisition et l'analyse de spectres (H)CC(CO)NH-TOCSY et H(CC)CONH-TOCSY, à partir desquelles chaque acide aminé est identifié par son groupement amide, autrement dit les fréquences du couple H<sub>N</sub>, N<sub>i</sub>. L'expérience (H)CC(CO)NH-TOCSY permet l'identification des carbones des chaînes latérales du résidu précédent. Une attribution spécifique de ces carbones est rendue possible par l'utilisation des gammes de déplacements chimiques présentés à la Figure 10 du chapitre précédent

Les plans <sup>13</sup>C-<sup>1</sup>H de l'expérience (H)CCH-TOCSY sont ensuite analysés et les systèmes de spins précédents identifiés. Les taches de corrélation observées permettent alors

pour chaque carbone d'un résidu quelconque, l'attribution des fréquences de résonances des protons correspondants.

Il est toutefois possible de déterminer les attributions des protons des chaînes latérales par l'enregistrement de l'expérience  $H(CCO)NH$ -TOCSY qui relie les fréquences de résonances des protons de la chaîne latérale  $i-1$  aux fréquences du couple  $H_{N_i,N_i}$ . Si elle se prête bien à une vérification, cette expérience est plus délicate pour une attribution en raison de sa faible résolution et sensibilité.

Si l'attribution de la chaîne principale de SufA sous sa forme apo a permis d'obtenir un score correct de 93.5% (voir page 75), il n'en a pas été de même avec les atomes des chaînes latérales et particulièrement pour les protons de chaînes aliphatiques longues (tableau ci-dessous). Ainsi suite aux premières analyses, il a été nécessaire d'enregistrer une nouvelle expérience de type  $HBHACONH$  pour parfaire l'attribution de ce type de protons. On a d'autre part moduler les paramètres des expériences RMN précédentes de manière à récupérer de l'information. Si nous avons réussi à améliorer l'attribution des chaînes latérales, le résultat obtenu est toutefois décevant et, comme nous le verrons, générera des difficultés dans le cadre de calcul de structure.

Notons que les expériences  $3D-^{15}N$ -NOESY-HSQC et  $3D-^{13}C$ -NOESY-HSQC enregistrées à 800 MHz nous ont permis également de vérifier l'attribution des résonances.

## -Bilan des attributions (Tableau 1).

AA	H <sub>a</sub>	H <sub>b</sub>	Autres	Nbre H th's.	Nbre H obs.	Ca	Cb	Autres	Nbre C th's.	Nbre C obs.
Ala 9	1	1	0	18	H <sub>a</sub> =9 H <sub>b</sub> =9 Autres=0 100%	1	1	0	18	Ca=8 Cb=8 Autres=0 100%
Arg 3	1	2	4	21	H <sub>a</sub> =1 H <sub>b</sub> =0 Autres=0 4.7%	1	1	2	12	Ca=3 Cb=3 Autres=5 51.5%
Asn 4	1	2	0	12	H <sub>a</sub> =3 H <sub>b</sub> =6 Autres=0 75%	1	1	0	8	Ca=4 Cb=3 Autres=0 87.5%
Asp 9	1	2	0	27	H <sub>a</sub> =7 H <sub>b</sub> =16 Autres=0 100%	1	1	0	18	Ca=9 Cb=9 Autres=0 100%
Cys 3	1	2	0	9	H <sub>a</sub> =3 H <sub>b</sub> =6 Autres=0 100%	1	1	0	6	Ca=3 Cb=3 Autres=0 100%
Gln 7	1	2	2	35	H <sub>a</sub> =6 H <sub>b</sub> =10 Autres=12 80%	1	1	1	21	Ca=7 Cb=7 Autres=6 95%
Glu 7	1	2	2	35	H <sub>a</sub> =4 H <sub>b</sub> =10 Autres=8 66%	1	1	1	21	Ca=7 Cb=7 Autres=6 95%
Gly 14	2	0	0	28	H <sub>a</sub> =22 H <sub>b</sub> =0 Autres=0 78%	1	0	0	14	Ca=14 Cb=0 Autres=0 100%
His 4	1	2	0	12	H <sub>a</sub> =4 H <sub>b</sub> =4 Autres=0 66%	1	1	0	8	Ca=4 Cb=4 Autres=0 100%
Ile 4	1	2	4	24	H <sub>a</sub> =3 H <sub>b</sub> =3 Autres=8 50%	1	1	3	20	Ca=4 Cb=3 Autres=5 60%
Leu 10	1	2	3	60	H <sub>a</sub> =9 H <sub>b</sub> =10 Autres=15 56%	1	1	3	50	Ca=8 Cb=8 Autres=16 64%
Lys 6	1	2	6	54	H <sub>a</sub> =5 H <sub>b</sub> =8 Autres=24 60%	1	1	3	30	Ca=6 Cb=6 Autres=12 80%
Met 4	1	2	2	20	H <sub>a</sub> =3 H <sub>b</sub> =4 Autres=4 55%	1	1	1	12	Ca=4 Cb=3 Autres=1 66%
Phe 10	1	2	0	30	H <sub>a</sub> =10 H <sub>b</sub> =10 Autres=0 66%	1	1	0	20	Ca=10 Cb=10 Autres= 100%
Pro 7	1	2	4	49	H <sub>a</sub> =6 H <sub>b</sub> =5 Autres=6 34%	1	1	2	28	Ca=6 Cb=6 Autres=9 75%
Ser 4	1	2	0	12	H <sub>a</sub> =4 H <sub>b</sub> =8 Autres=0 100%	1	1	0	8	Ca=4 Cb=4 Autres=0 100%
Thr 5	1	2	1	15	H <sub>a</sub> =3 H <sub>b</sub> =4 Autres=3 66%	1	1	1	15	Ca=5 Cb=5 Autres=4 93%
Trp 1	1	2	0	3	H <sub>a</sub> =0 H <sub>b</sub> =0 Autres=0	1	1	0	2	Ca=1 Cb=1 Autres=0 100%
Tyr 1	1	2	0	3	H <sub>a</sub> =0 H <sub>b</sub> =0 Autres=0	1	1	0	2	Ca=1 Cb=1 Autres=0 100%
Val 10	1	2	2	40	H <sub>a</sub> =8 H <sub>b</sub> =7 Autres=16 77%	1	1	2	40	Ca=10 Cb=10 Autres=10 95%
				507	327 64.5%				353	301 85%

Tableau 1: Bilan des attributions des <sup>1</sup>H et <sup>13</sup>C des chaînes latérales de l'ensemble des résidus de SufA

Dans la colonne des noms des acides aminés est précisé le nombre trouvé dans la séquence de SufA. D'autre part dans ce décompte, on ne tient pas compte des cycles des résidus aromatiques qui ont fait l'objet d'expériences particulières. Le premier constat que l'on peut faire quand on analyse ce tableau est que le nombre de protons attribués des chaînes latérales n'est pas des plus élevé même si cela reste raisonnable (64.5%). On ne trouve rien de systématique qui puisse expliquer ces manques. Rien qui puisse nous faire dire pourquoi telle ou telle zone de la protéine est moins attribuée. Cette perte d'information touche l'ensemble de la protéine mais on constate toutefois que les attributions des protons H<sub>a</sub> et H<sub>b</sub> sont relativement mauvaises pour les résidus aromatiques et, d'autre part, que les chaînes aliphatiques longues (Arg, Ile, Leu) semblent souffrir d'un manque d'attribution tant au niveau des protons que des carbones. Cela touche aussi bien les résidus hydrophobes (Ile par exemple) qu'hydrophiles (Arg par exemple).

Une explication probable pour cette perte d'information est la taille de la protéine dont la valeur du temps de corrélation  $\tau_c$  induirait un T<sub>2</sub> plus court ne favorisant pas ainsi le



transfert d'aimantation au cours des expériences triple résonance. D'autre part des mouvements locaux peuvent aussi jouer sur la relaxation de ces systèmes de spins, avec la même conséquence.

Une expérience HBHACONH spécifique pour les protons  $H\alpha$  et  $H\beta$  a été enregistrée mais elle ne nous a pas permis d'augmenter le score des attributions protons.

### III.6-Attributions des aromatiques

Ces résidus représentent 13% du nombre total d'acides aminés ce qui n'est pas négligeable. Des molécules telles que la phénylalanine et le tryptophane sont hydrophobes alors que l'histidine est hydrophile. La tyrosine, quant à elle est neutre. Ce qui est intéressant pour notre structure, c'est qu'il y a dix phénylalanines qui auront tendance à être dans des zones hydrophobes. Par conséquent, elles sont enfouies dans la structure et on devrait s'attendre à mesurer des effets nOe facilement identifiable. La difficulté d'attribution pour ce type de résidu, réside par contre dans le risque de superposition que nous pourrions rencontrer (vu le nombre de phénylalanines).

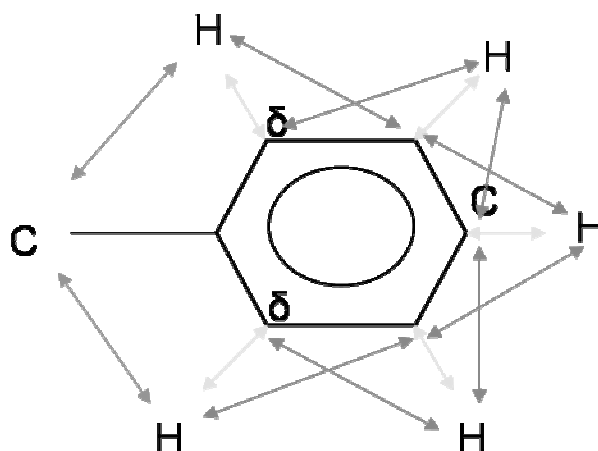
Pour l'attribution des résidus aromatiques nous avons dans un premier temps enregistré une série d'expériences 2D spécifiques.

$^{13}\text{C}$  HSQC – AROM qui doit permettre d'identifier les carbones du cycles et leur proton respectif.

2D (H)C(C)H TOCSY – AROM qui doit permettre de connecter un proton avec les carbone voisins  $i-1$  ou  $i+1$ .

2D CBHD – AROM qui doit connecter le  $C\beta$  d'une chaîne latérale au  $H\delta$  du cycle aromatique permettant ainsi de connecter les attributions aromatiques à celles de la chaîne latérale et ainsi identifier le résidu aromatique concerné.

On peut schématiser comme suit



en vert : 2D CBHD – AROM

en orange : 2D (H)C(C)H TOCSY

en jaune :  $^{13}\text{C}$  HSQC – AROM

-Bilan des attributions (Tableau 2)

			$\text{H}^{\delta 2}$	$\text{C}^{\delta 2}$			$\text{H}^{\epsilon 1}$	$\text{C}^{\epsilon 1}$		
His4			0	0			0	0		
His27			1	1			0	0		
His72			1	1			1	1		
His106			1	1			1	1		
	$\text{H}^{\delta 1}$	$\text{C}^{\delta 1}$	$\text{H}^{\delta 2}$	$\text{C}^{\delta 2}$	$\text{H}^{\epsilon 1}$	$\text{C}^{\epsilon 1}$	$\text{H}^{\epsilon 2}$	$\text{C}^{\epsilon 2}$	$\text{H}^{\zeta}$	$\text{C}^{\zeta}$
Phe8	1	1	0	0	1	1	0	0	1	1
Phe13	1	1	0	0	1	1	0	0	0	0
Phe53	1	1	0	0	1	1	0	0	1	1
Phe70	1	1	0	0	0	0	0	0	0	0
Phe78	1	1	0	0	0	0	0	0	0	0
Phe86			0	0	0	0	0	0	0	0
Phe94	1	1	0	0	1	1	0	0	1	1
Phe103	1	1	0	0	1	1	0	0	1	1
Phe105	1	1	0	0	1	1	0	0	1	1
Phe120	0	0	0	0	0	0	0	0	0	0

	H <sup>δ1</sup>	C <sup>δ1</sup>	H <sup>ε2</sup>	C <sup>ε2</sup>	H <sup>ε3</sup>	C <sup>ε3</sup>	H <sup>ε2</sup>	C <sup>ε2</sup>	H <sup>η2</sup>	C <sup>η2</sup>
Trp15	1	1	1	1	1	1	1	1	1	1
	H <sup>δ1</sup>	C <sup>δ1</sup>	H <sup>δ2</sup>	C <sup>δ2</sup>	H <sup>ε1</sup>	C <sup>ε1</sup>	H <sup>ε2</sup>	C <sup>ε2</sup>		
Tyr55	0		0	0	0	0	0	0		

Tableau 2 : Bilan des attributions des résidus aromatiques.

On décompte 29 protons attribués sur 67 attendus, de même pour les carbones, ce qui représente 43% d'attribution pour les aromatiques. Le bilan de cette attribution mitigé est du aux nombreuses superpositions des résonances dans les spectres 2D ainsi que du manque d'attribution au niveau des H $\alpha$  et H $\beta$ . Ceci empêche de connecter le cycle à la chaîne latérale. De plus la mauvaise sensibilité de l'expérience TOCSY n'a pas favorisé l'attribution de ces résidus.

Une solution pour pallier à ces problèmes d'attributions serait peut être de deutérer la protéine, mais cela n'a pas été l'option choisie, et à ce stade de l'étude dans le cadre de cette thèse cela ne semblait pas complètement opportun.

### III.7-Note d'attribution

L'attribution des résonances <sup>1</sup>H, <sup>15</sup>N, <sup>13</sup>C de SufA a fait l'objet d'une publication dans Journal of biomolecular NMR. La liste des déplacements chimiques a été déposée dans la base de donnée BioMagResBank (<http://www.bmbr.wisc.edu>) avec l'identifiant BMRB-6224.

### III.8-Échange de protons amides

Le troisième type d'information qui confirme la localisation des éléments de structures secondaires est la vitesse d'échange du proton amidique avec le deutérium. Cette mesure de la vitesse d'échange permet d'identifier les zones de la protéine accessibles au solvant. En effet, un résidu enfouit au cœur de la protéine, ou impliqué dans une liaison hydrogène dans les structures secondaires, verra son proton H<sub>N</sub> s'échanger très lentement, ou pas, avec le deutéron de l'eau lourde.

Nous avons déterminé le rapport d'échange H/D,  $k_{ex}$ , pour SufA. Pour ce faire, un échantillon de protéine SufA simplement marquée  $^{15}\text{N}$  a été lyophilisé et dissout dans de l'eau lourde de pH égale à 7. Une série d'expériences  $^1\text{H}$ - $^{15}\text{N}$  HSQC a été ensuite enregistrées à intervalles de temps réguliers sur une durée définie. L'idée est de suivre la décroissance de l'intensité du signal de corrélation  $^1\text{H}$ - $^{15}\text{N}$  en fonction du temps. Des courbes obtenues ( $I = f(T)$ ), on détermine la vitesse d'échange, en fait le rapport d'échange  $k_{ex}$ , en tenant compte de constantes physico-chimiques tels que les pKa des résidus et le pH auquel on opère.

L'enregistrement des spectres s'est fait sur une durée de 17 heures à intervalles de temps croissant régulièrement (période courte au début et longue à la fin). Les intensités ont été mesurées en fonction du temps pour chacun des résidus de la protéine et les courbes obtenues ont été interpolées afin d'en extraire le rapport d'échange  $k_{ex}$ . Ce rapport n'est pas exploité de manière brute mais est comparé à un rapport intrinsèque,  $k_{int}$  qui prend en compte l'environnement stérique et chimique du proton considéré ( $\log(k_{int}/k_{ex})$ )(Figure 10). L'ensemble de ces paramètres, ainsi que la méthodologie de traitement est présentée dans le chapitre « matériels et méthodes », dans la partie expériences RMN.

Nous obtenons les résultats suivants :

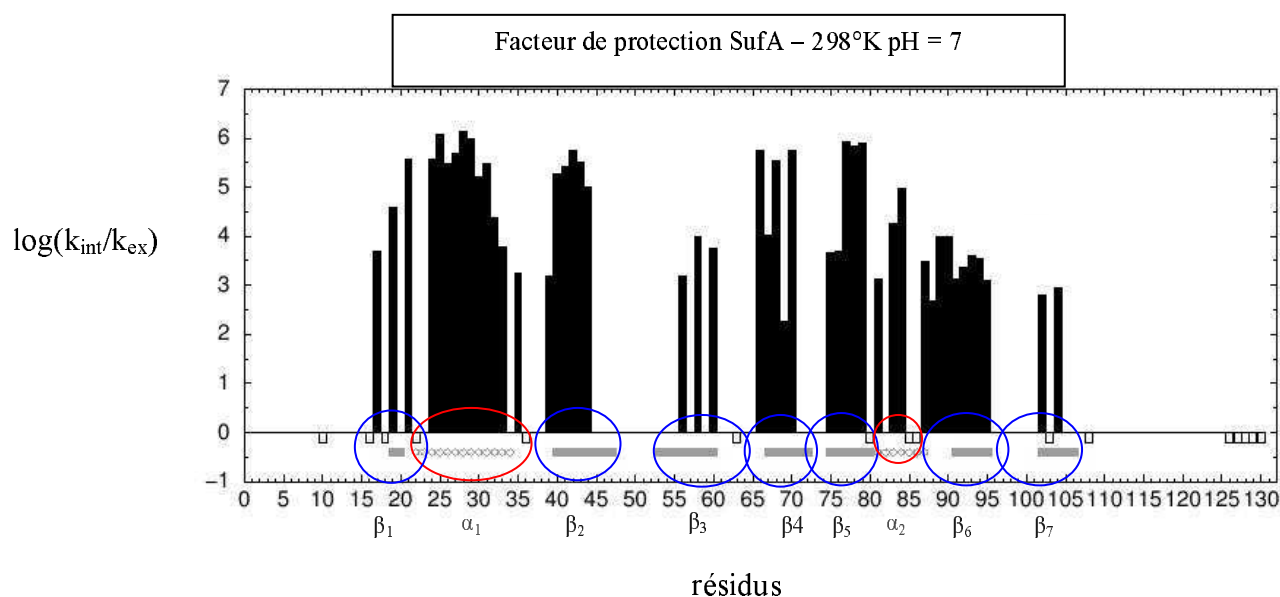


Figure 10 : On constate que l'accessibilité des protons échangeables est faible pour les résidus impliqués dans les structures secondaires. En effet ces protons pourraient être engagés dans des liaisons hydrogènes qui ne favorisent pas l'échange.

On constate que globalement, les protons pour lesquels le facteur de protection est significatif, se situent dans la plupart des résidus que l'on a identifiés comme faisant partie de structures secondaires. Par contre, au niveau des résidus 83, 84 et 85 il semblerait qu'il existe une structure secondaire.

## Chapitre IV

### Détermination de la structure de SufA

La structure d'une protéine en solution est déterminée par un calcul de modélisation moléculaire prenant en compte deux types de données. D'une part, des données disponibles a priori (longueurs de liaison covalente, angle de valence, planéité, rayons de van der Waals, etc) constitue le « champ de force ». D'autre part, des données extraites d'expériences de Résonance Magnétique Nucléaire (déplacements chimiques, contacts nOe, couplages scalaires, couplages dipolaires résiduels, etc) constituent un « jeu de contraintes » (liaisons interprotons, angles dièdres, orientations à longue portée, etc). La conformité d'un modèle moléculaire avec ces deux types de données est estimée grâce à une « fonction d'énergie » hybride que l'on cherche à minimiser de manière à déterminer la structure de plus basse énergie. Tout le calcul de structure de la protéine SufA a été faite avec le logiciel ARIA qui suit cette procédure de calcul.

La stratégie que nous avons adoptée a été de calculer tout d'abord la structure du monomère ; et c'est ce calcul qui est présenté dans ce qui suit. Dans un second temps, une fois la structure du monomère validée, nous tenterons de construire un modèle du dimère à partir de la structure du monomère et de contraintes nOe intermoléculaires que nous avons extraites des expériences RMN. Une des difficultés a d'ailleurs été de discriminer ce qui était intermoléculaire et intramoléculaire.

La détermination de la structure de la protéine SufA a nécessité la collecte de contraintes structurales de distances par la mesure de corrélations nOe à partir d'expériences 3D-NOESY HSQC éditées  $^{13}\text{C}$  et  $^{15}\text{N}$ , de contraintes orientationnelles avec la mesure de couplages dipolaires résiduels par des expériences 3D hétéronucléaire de type HNCO. Notons que la protéine SufA est dimère et que le calcul présenté ci-après a été fait pour le monomère.

## I-CONTRAINTES DE DISTANCES

Une expérience 3D-NOESY HSQC éditée  $^{13}\text{C}$  ainsi qu'une expérience éditée  $^{15}\text{N}$  ont été enregistrées pour déterminer les contraintes structurales de distances à partir de la mesure de corrélations nOe ainsi que deux expériences de même type, mais spécifiques aux résidus aromatiques. Une expérience 3D NOESY-HSQC pour la mesure de distance  $\text{CH}_3\text{-CH}_3$  [ $^{136}$ ] a aussi été enregistrée. Le calcul de structure a été effectué à partir de 1652 nOe issus de l'expérience NOESY éditée  $^{13}\text{C}$ , de 824 nOe issus de l'expérience NOESY  $^{15}\text{N}$ , de 59 nOe issus des expériences NOESY spécifiques aux résidus aromatiques ainsi que 62 nOe issus de l'expérience filtrée  $\text{CH}_3$ . On obtient donc un nombre important de contraintes à gérer.

Comme nous le verrons plus loin cette tâche peut être facilitée par l'utilisation du logiciel ARIA qui effectue une attribution des contraintes nOe au cours des calculs. Nous verrons que plus les calculs convergent vers une structure stable, plus les informations structurales issues du repliement de la protéine permettent de lever les ambiguïtés d'attribution (pics pouvant être attribués à plusieurs paires d'atomes). Et l'augmentation du nombre de nOe attribués permet d'obtenir des structures mieux définies. On comprend bien ici que la stratégie du calcul réside en un aller retour permanent, du moins dans une première phase, entre le calcul de modélisation moléculaire sous contraintes et l'attribution des spectres NOESY. Ceci est totalement vérifié dans le cas de protéines dont l'attribution est pratiquement complète, c'est-à-dire que l'ensemble des atomes qui constituent la protéine est marqué par une valeur de déplacement chimique. En effet, dans ces conditions le niveau d'ambiguïté est moindre et l'attribution des spectres nOe s'en trouve facilitée. En ce qui concerne la protéine SufA, telle n'est pas la situation au moment où nous débutons l'attribution des pics nOe. En effet, un pourcentage non négligeable des protons des chaînes latérales (35%), ainsi que de carbones (15%), n'est pas attribué.

On distingue classiquement 4 à 5 types de contacts nOe [ $^{137}$ ] : les nOe intrarésiduels (entre protons appartenant au même résidu), les nOe séquentiels (entre protons appartenant à des résidus voisins dans la séquence), les nOe à moyenne portée (entre protons appartenant à des résidus distants de 2 à 4 résidus dans la séquence), les nOe à longues portées (entre protons appartenant à des résidus de 5 résidus et plus dans la séquence) et, le cas échéant, les nOe intermoléculaires (entre protons de macromolécules distinctes). Cette classification est

uniquement basée sur la position relative des protons dans la séquence et non pas dans l'espace. Les nOe intrarésiduels et séquentiels sont les plus faciles à identifier mais ils n'apportent pas d'information sur la structure tertiaire de la protéine. Ils sont utiles lors de l'étape d'attribution séquentielle des résonances et donnent une indication de la structure secondaire (proximité du proton amide  $i$  avec le proton amide  $i+1$  dans les hélices  $\alpha$  ; proximité du proton  $\alpha_i$  avec le proton  $H_{i+1}^N$  dans les brins  $\beta$ ). Les nOe à moyenne distance sont utiles dans la définition de la structure secondaire de la protéine, en particulier dans les hélices  $\alpha$  (proximité du proton  $H_{i+1}^N$  avec le proton  $H_{i+3}^N$  et dans une moindre mesure avec le  $i+2$  et  $i+4$ ). Les nOe à longue portée sont essentiels pour la définition de la structure tertiaire de la protéine.

### I.1-Extraction des nOe courtes portées

La détermination de ces contacts nOe intra-résidus et séquentiels a été effectuée grâce à un programme écrit au laboratoire RMN qui déduit l'emplacement des pics intrarésiduels et séquentiels à partir du fichier de déplacement chimique au format STAR (BMRB). Il s'agit de scripts PERL, `Noehsqc_ppm_maker_c13.pl` pour l'expérience 3D NOESY HSQC  $^1H$ - $^{13}C$  et `Noehsqc_ppm_maker_n15.pl`, pour l'expérience 3D NOESY HSQC  $^1H$ - $^{15}N$ . Si l'attribution de ces contacts permet de mieux identifier les contraintes longue portée, elle est très utile dans la détermination de la conformation locale de la protéine.

### I.2-Extraction des nOe inter-brins $\beta$ et moyenne portée

L'expérience 3D  $^{15}N$ -NOESY-HSQC enregistrée à 800 MHz est utilisée pour extraire les nOe séquentiels et intra-résiduels. Cette expérience permet de délimiter et d'identifier les éléments de structures secondaires avec les nOe séquentiels puis de reconstituer la topologie des feuillets  $\beta$  en cherchant des nOe  $H_N$ - $H_N$  et  $H_N$ - $H_\alpha$  entre les brins. Dans les hélices  $\alpha$ , un nOe très intense est observé entre le proton amide d'un résidu ( $i$ ) avec le proton amide du résidu précédent ( $i-1$ ) ou avec le proton  $\alpha$  du résidu ( $i-3$ ). Pour les brins  $\beta$ , le nOe le plus intense vient du contact du proton amide d'un résidu ( $i$ ) avec le proton  $\alpha$  du résidu précédent (Figure 1). Si on reprend la figure de la note d'attribution, on observe toujours une bonne corrélation entre les différents moyens d'identification des éléments de structures secondaires.



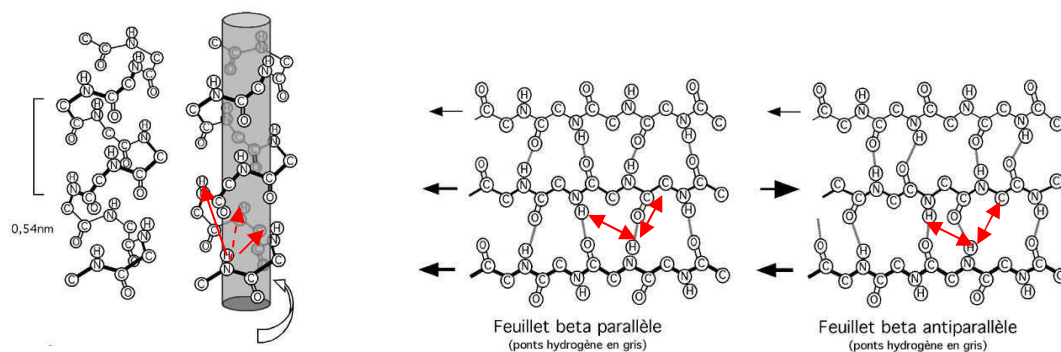


Figure 1 : nOe séquentiels 1H-1H dans les éléments de structures secondaires

L'attribution de ces corrélations permet de définir la topologie des feuillets  $\beta$  ainsi que celle de l'hélice  $\alpha$ . En ce qui concerne la protéine SufA, nous avons pu identifier les structures secondaires (exemple Figure 2) qui correspondent à celles prévues par l'analyse du CSI (comme de la prédiction de structure secondaire). Nous confirmons donc bien la présence de 7 brins  $\beta$  qui semblent constituer deux feuillets et une hélice  $\alpha$ . A ce stade, nous n'avons toujours pas pu identifier une seconde (et petite) hélice  $\alpha$  qui apparaît dans la structure de la protéine homologue IscA.

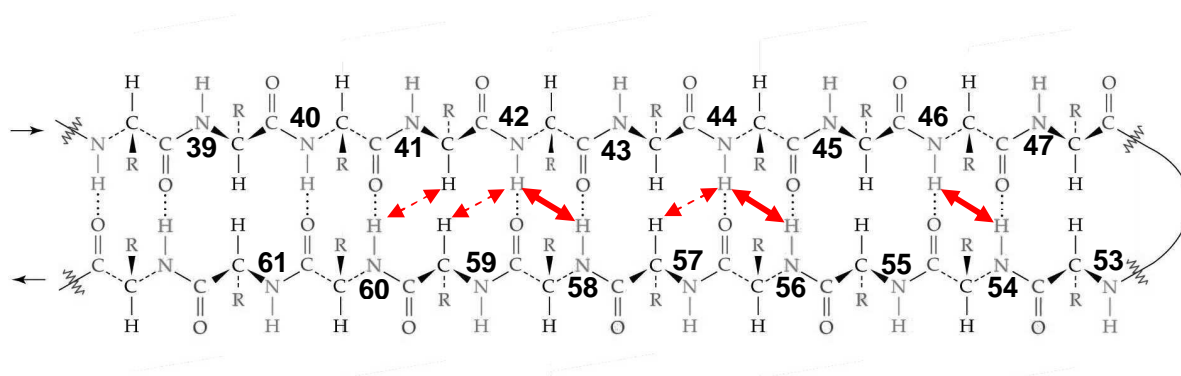


Figure 2: Exemple de contraintes nOe interbrins  $\beta$  obtenues sur SufA. Les flèches en rouges épais correspondent à des nOe intenses alors que celles en pointillées correspondent à des nOe de faible intensité.

L'ensemble de ces mesures est très bien résumé dans le schéma de la note d'attribution représenté ci-après sur la Figure 3.

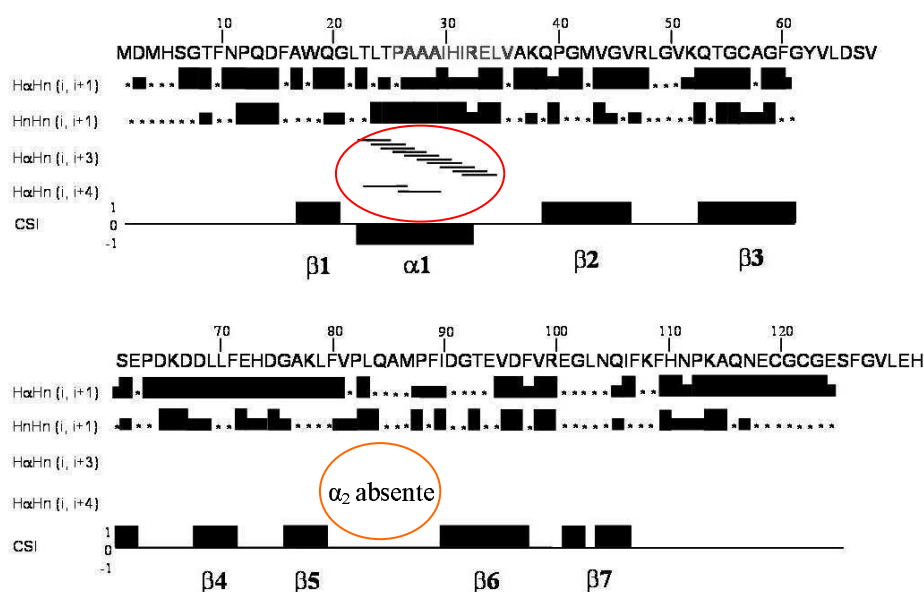


Figure 3: Les nOe séquentiels et intrarésidus sont indiqués par les barres noires dont la hauteur indique l'intensité.

\* sont les nOe non observés.

On observe parfaitement l'hélice  $\alpha_1$  qui présente bien contacts nOe  $H_aH^N(i, i+3)$  et  $(i, i+4)$ .

### I.3-Extraction des nOe longue portée

Ces nOe ont certainement la plus grande valeur structurale. On pourrait essayer d'attribuer manuellement l'ensemble des corrélations nOe mais ce travail s'avère très fastidieux avec un résultat incertain. En fait, la plupart des stratégies de calcul sont basées sur l'utilisation d'un nombre réduit de contraintes à longue portée et non ambiguës pour déterminer le repliement global de la protéine. Cette structure globale sert alors de base aux calculs suivants, au cours desquels l'attribution (donc l'extraction) des nOe longue portée est effectuée. Pour notre part nous avons opté pour l'utilisation du logiciel qui est un algorithme de calcul de structure qui permet une attribution semi automatique des contraintes nOe ambiguës.

Plusieurs facteurs peuvent rendre une attribution ambiguë. Tout d'abord, des résonances ayant le même déplacement chimique peuvent se superposer. Ce déplacement chimique a d'ailleurs une précision limitée par la résolution spectrale (largeur de raie) et digitale (nombre de points) de spectre et par les chevauchements de pics qui décalent les résonances apparentes. D'autre part les conditions d'enregistrement (pH, température) de différents spectres peuvent influencer la valeur de déplacement chimique d'un même atome. La levée de l'ambiguïté se fait dans toutes les approches actuelles via un processus itératif au

cours duquel la structure calculée permet la levée des ambiguïtés qui, à leur tour, sont injectées dans le calcul suivant pour lequel la convergence s'améliore. Cette amélioration doit ensuite aider à lever de nouvelles ambiguïtés qui de nouveau sont injectées dans le prochain calcul et ainsi de suite. En résumé la philosophie du calcul est qu'à chaque nouvelle étape de calcul les nouvelles contraintes permettent de définir un peu plus précisément la structure et chaque structure ainsi calculée permet de lever des ambiguïtés sur les contraintes avec une confiance de plus en plus élevée. Il est bien évident que l'étape fondamentale est l'étape initiale où on ne doit pas propager d'erreurs d'attribution.

L'ensemble de ces contraintes de distances ont été extraites suite à l'analyse des spectres RMN avec le logiciel NMRview [99] après traitement des fid (enregistrées sur les spectromètres VARIAN de 600 et 800 MHz) par NMRpipe<sup>[138]</sup>.

Les corrélations nOe ont été obtenues par l'analyse des deux spectres <sup>15</sup>N-NOESY-HSQC et <sup>13</sup>C-NOESY-HSQC pour lesquels un certain nombre de corrélations nOe, les courtes et moyennes portées, étaient déjà attribuées via les protocoles décrits ci-dessus. Le spectre <sup>15</sup>N-NOESY-HSQC a été ensuite attribué manuellement, tout comme le spectre <sup>13</sup>C-NOESY-HSQC dans un premier temps. En effet, si ARIA est conçu pour gérer les ambiguïtés, il est toutefois intéressant d'effectuer une attribution partielle, dont on est sûr, des spectres de manière à aider le logiciel à replier la protéine.

Les autres contraintes longues portées utilisées, sont celles issues d'une expérience <sup>13</sup>C NOESY-HSQC spécifique qui permet de ne mesurer que les effets nOe entre résidus portant un groupement méthyle (CH<sub>3</sub>). L'idée est d'obtenir ainsi un nombre suffisant de contraintes nOe non ambiguës longues distances, entre ces méthyles, qui doit aider au repliement de la protéine dans les premiers calculs d'une structure basse résolution.

Les dernières expériences que nous avons enregistrées sont des expériences <sup>15</sup>N-NOESY-HSQC et <sup>13</sup>C-NOESY-HSQC spécifiques aux résidus aromatiques. (cf chapitre matériels et méthodes). L'intérêt de ces résidus réside dans le fait qu'ils sont souvent localisés dans les poches hydrophobes des structures et sont par conséquent dans des zones relativement structurées. Ils permettent d'obtenir des contraintes nOe longue portée plus facilement identifiables, vu les déplacements chimiques particuliers des atomes de <sup>13</sup>C et <sup>1</sup>H des cycles aromatiques. En ce qui concerne la protéine SufA, l'analyse des expériences propres à ces résidus a présenté des difficultés d'attribution. Principalement pour deux raisons, premièrement le nombre important de phénylalanines (10) pour lesquels le

recouvrement spectral était très important n'a pas permis une attribution exhaustive de ces résidus. D'autre part les résidus Tyrosine et Tryptophane ne sont pas attribués. Par conséquent, l'attribution des contacts nOe des spectres NOESY-HSQC des résidus aromatiques n'a été que partielle. Les systèmes identifiés nous ont tout de même permis d'augmenter le nombre de pics nOe attribués.

Le déficit de protons attribués au niveau des chaînes latérales et le fait que nous soyons en présence d'un dimère nous a incité à suivre une stratégie particulière pour la détermination de la structure de la protéine SufA. En effet, nous avons décidé d'effectuer deux calculs en parallèle. Un premier calcul sans structure de base c'est-à-dire à partir de la protéine linéaire et un deuxième calcul avec une structure de base. Celle-ci est la structure de la protéine homologue IscA que nous injectons donc dans le calcul dans le répertoire initial. Chacun des calculs a été effectué à partir des mêmes fichiers de pics (fichiers xpk) issus de toutes les expériences RMN traitées et analysées avec le logiciel de visualisation NMRView. Dans les deux cas, a été ajouté un fichier de contraintes TALOS (voir paragraphe II) (fichier talos.tbl) issu de l'analyse des déplacements chimiques suite à l'attribution des atomes. Notons qu'un certain nombre de contraintes nOe des différentes expériences ont été attribuées manuellement et nous avons fait le choix de ne pas laisser la possibilité au calcul de structure de réattribuer ces contraintes. En effet, l'attribution manuelle a été réalisée sans ambiguïté sur les systèmes de spins concernés.

A l'issue des calculs avec et sans modèle structural de départ, nous avons fusionné les nOe qui présentaient la même attribution dans chacun des calculs en éliminant les redondances. Le nombre de contraintes nOe obtenu est présenté dans le tableau 1 dans la colonne C.

A ce stade nous avons mis de côté les candidats susceptibles d'être en interaction intermoléculaire. Ces contraintes intermoléculaires ont été déterminées sur la base de structure impossible à obtenir dans le cadre du calcul du monomère. D'autre part, sur la base d'une des deux structures obtenues, nous avons déterminé le tenseur d'alignement de la protéine avec le logiciel « Module » à partir des fichiers de couplages dipolaires résiduels.

Notons que la liste de contraintes nOe issue des deux calculs a permis des nouvelles analyses des spectres RMN afin d'ajouter ou de corriger les attributions.

Les calculs suivants ont été ensuite tous effectués à partir d'une structure linéaire de SufA, afin de s'affranchir de la structure cristallographique et par conséquent de ne déterminer le calcul de structure que sur la base des données expérimentales RMN. Pour tous

les calculs qui ont suivi, on a introduit l'ensemble des contraintes dont nous disposions, soit le fichier de contraintes talos.tbl, les fichiers de contraintes orientationnelles rdc.tbl, les contraintes de distances attribuées ambig.tbl et unambig.tbl, le fichier des liaisons hydrogène sous forme de contraintes de distances H-bond.tbl et les fichiers de pics des expériences RMN file.xpk. Ceci nous a permis d'aboutir à un calcul avec une bonne convergence des structures de plus basses énergies que nous analyserons ci-après. Le tableau suivant indique les résultats obtenus quant à l'analyse des contraintes nOe au cours des différents calculs (en fait pour le calcul initial, pour le calcul intermédiaire issu des deux stratégies et pour le calcul final) pour les différents spectres RMN enregistrés.

#### -Bilan des attributions

SPECTRE	A) Nombre de pics	B) Attribués manuellement	C) Attribués après 1 <sup>ère</sup> étape d'Aria non-ambiguës	D) Attribués total	E) Attribués ambiguës
<sup>13</sup> C NOESY HSQC	2519	1399 (1138) 627 intra 202 seq 110 med 191 LR	1765 (1434) 754 intra 257 seq 141 med 282 LR	2129 (1573) 790 intra 283 seq 148 med 352 LR	79
<sup>15</sup> N NOESY HSQC	935	754 (711) 311 intra 317 seq 41 med 42 LR	829 (779) 339 intra 333 seq 51 med 56 LR	883 (807) 344 intra 341 seq 54 med 68 LR	17
<sup>13</sup> CH <sub>3</sub> NOESY HSQC	101	82 (56) 20 intra -seq 6 med 30 LR	101 (59) 20 intra - 6 med 33 LR	101 (59) 20 intra - 6 med 33 LR	
AROM <sup>13</sup> C NOESY HSQC-	85	51 (45) 37 intra 1 seq - 7 LR		82 (61) 37 intra 2 seq - 22 LR	1

Les nombres entre parenthèses correspondent aux attributions de pics nOe uniques (c'est-à-dire sans doublon). On constate que le nombre de contraintes attribuées augmente bien au cours des différents calculs au fur et à mesure que la structure converge comme attendu. En clair, lorsque le niveau d'ambiguïté diminue les structures convergent mieux. Ce constat empirique, presque évident, doit toutefois être accompagné d'autres critères de jugement tel que le nombre de violations des contraintes de distances, et des autres contraintes par ailleurs.

## II-CONSTRAINTES D'ANGLES DIÈDRES

### -TALOS (Torsion Angle Likelihood Obtained from Shift and sequence similarity):

Les angles  $\varphi$  et  $\psi$  ont été estimés pour chaque résidu de SufA avec le système de banques de données de TALOS [<sup>139</sup>]. Les déplacements chimiques des atomes  $H_\alpha$ ,  $C_\alpha$ ,  $C_\beta$ , CO, N ont été utilisés et la comparaison des déplacements chimiques s'est effectuée avec les 20 protéines dans la banque de données de TALOS. La contrainte est acceptée si la corrélation est bonne dans au moins 9 cas sur 10. En ce qui concerne les contraintes TALOS nous n'avons gardé que celles qui nous avaient donné un score de 9 ou 10. D'autre part, nous avons fait le choix arbitraire de ne conserver que les contraintes angulaires des structures secondaires. Ainsi 332 contraintes TALOS sont utilisées pour les calculs.

## III-CONSTRAINTES ORIENTATIONNELLES

Les couplages dipolaires résiduels comme nous l'avons vu dans le chapitre 2, fournissent des informations structurales qui vont nous permettre d'affiner la structure de la protéine SufA. Pour ce faire, nous avons mesuré les couplages dipolaires résiduels dans un milieu orientant qui est une solution de bactériophages *Pf1*. Dans ce milieu les mécanismes d'alignement sont dominés par des interactions électrostatiques.

### III.1-Préparation du milieu cristal liquide et des échantillons

#### III.1.1-SufA dans les bactériophages *Pf1*

Les bactériophages *Pf1* utilisés ont été achetés chez Asla Ltd. Les phages à 50 mg/mL sont dans du tampon phosphate 10 mM pH 7.6 avec 2 mM de  $\text{MgCl}_2$  et 0.05% de  $\text{NaN}_3$ . Il est nécessaire de les conserver à 4°C. Pour les utiliser, on peut les désagréger en prélevant la quantité nécessaire pour l'échantillon à préparer puis en les mettant dans le milieu dans lequel est la protéine, à savoir un tampon phosphate 50 mM à pH 6.9 en présence de DTT, de Complete<sup>TM</sup> et de 10% en volume final de  $\text{D}_2\text{O}$ . Nous avons utilisé 50 mg/mL de phages pour le premier échantillon pour lequel on a mesuré le couplage quadripolaire du deutérium afin de déterminer le degré d'alignement du milieu. Le dédoublement de la raie deutérium était de 39.9Hz ce qui est une valeur normale dans la gamme des couplages quadripolaires résiduels de cet isotope en fonction de la concentration en phages établie par Hansen et *al* [126]. Suite à ce premier contrôle, nous avons effectué la même mesure mais en présence de la protéine doublement marquée  $^{15}\text{N}$  et  $^{13}\text{C}$  à 1mM. Le dédoublement était plus faible 17.7 Hz mais suffisant pour poursuivre l'étude. Nous avons enregistré un spectre  $^1\text{H}$ - $^{15}\text{N}$  HSQC sur cet échantillon (Figure4). On constate que l'empreinte de la protéine pour ce type de spectre est la même que dans le milieu liquide. On peut donc considérer que la protéine est restée intègre et ne colle pas au phages dans le milieu orientant et que l'on peut passer à l'étape suivante qui consiste à enregistrer les expériences nécessaires à la détermination des couplages dipolaires résiduels.

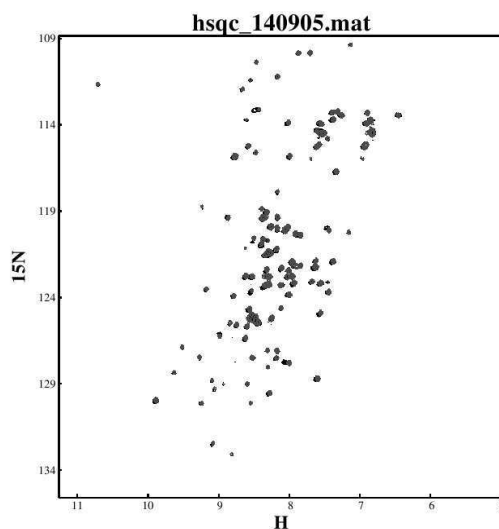


Figure 4 : Superposition de spectres  $^1\text{H}$ - $^{15}\text{N}$  HSQC de SufA 1mM pH7 tampon phosphate 50mM. En rouge SufA en milieu orientant anisotrope, en noir en milieu liquide isotrope.

On observe que la protéine dans l'échantillon présente le même repliement.

### III.1.2-Mesure de couplages dipolaires résiduels le long de la chaîne polypeptidique

#### III.1.2.1-Principe de la mesure

Le but est de mesurer un couplage  $J_{IS}^{\text{eff}}$  entre deux spins I et S qui induit un dédoublement de la raie de résonance dans un spectre RMN. Afin de résoudre les superpositions spectrales pour les protéine de grande taille, on peut mesurer ces couplages avec des expériences 3D de type HNCO.

Lorsque le couplage est plus petit que les largeurs de raies des composantes du doublet et pour palier le problème de superposition spectrale des protéines de grandes tailles, différents types de solutions ont été proposés dans la littérature, notamment trois familles d'expériences.

La première est la spectroscopie accordéon. Cette méthode consiste à séparer les composantes du doublet par un facteur  $\lambda J$  dans la dimension  $^{13}\text{C}$  en faisant évoluer le couplage durant une période plus longue ( $\lambda t_1$ ) que celle de l'évolution du déplacement chimique ( $t_1$ ) [140].

La deuxième est la spectroscopie sélective à l'état de spin passif S. Cette méthode consiste à séparer les composantes du doublet dans deux spectres différents, ce qui peut être une alternative à la première méthode si elle ne suffit à séparer les deux composantes du doublet quand le couplage est trop faible. Les valeurs de couplages se mesurent par la différence des fréquences des pics dans les deux spectres. Les trois séquences RMN utilisées sont :  $\text{S}^3\text{E}$  [141], IPAP [142] et DIPSAP [143]. Avec la séquence IPAP, le couplage évolue d'abord en phase puis en antiphase. L'addition et la soustraction de ces spectres conduit à deux spectres contenant chacun une composante du doublet. La séquence DIPSAP permet le même résultat mais avec une séparation plus propre des deux spectres, étant moins sensible que la séquence IPAP à la variation du couplage.

La troisième méthode est la spectroscopie E-COSY [141]. Cette méthode consiste à introduire un second couplage de grande amplitude dans une autre dimension pour séparer les pics de corrélation.



### III.1.2.2-Expériences enregistrées sur SufA

Nous avons mesuré trois types de couplages résiduels dipolaires :  $^1D_{NH}$ ,  $^2D_{COH}$  et  $^1D_{COCA}$ . Ils ont été mesurés sur l'échantillon doublement marqué, tout d'abord en milieu isotrope (identique à celui utilisé pour les attributions) puis en milieu anisotrope (phages *Pfl*), sur la base d'expériences 3D de type HN(CO)CA. La spectroscopie accordéon a été utilisée pour le couplage  $^1D_{NH}$ , E-COSY pour le couplage  $^1D_{COCA}$  alors que le couplage  $^1D_{CON}$  a été mesuré via la séquence DIPSAS.

### III.1.2.3-Extraction des couplages dipolaires résiduels

Les constantes de couplages dipolaires résiduels sont extraites en faisant la différence des couplages J+D mesuré en milieu anisotrope et J en milieu isotrope. Le bilan des couplages extraits est présenté dans le tableau 1.

Couplages	Gamme (Hz)	Nombre	%
$^1D_{NH}$	[-11.7 ; 15.2]	101	77
$^1D_{CON}$	[-5.6 ; 5.5]	94	72
$^1D_{COCA}$	[-7.5 ; 3.6]	99	76
<b>Total</b>		<b>294</b>	<b>75</b>

Tableau 1 : Gammes et nombres de RDC mesurés dans les bactériophages *Pfl*

Cet ensemble de données permet de créer un fichier de contraintes supplémentaires pour le calcul de structure de la protéine SufA comme nous l'avons déjà indiqué. Mais pour une utilisation optimale de ces contraintes orientationnelles, il est nécessaire d'estimer les valeurs propres du tenseur d'alignement. Pour cela, nous avons injecté les valeurs des RDC dans le logiciel MODULE [144] qui, sur la base d'une structure de SufA calculée avec les contraintes nOe, calcule les valeurs propres du tenseur.

Les valeurs obtenues et qui sont utilisées pour le calcul de structure sont :

$$R = 0.349$$

$$D_{HN} = +7.27$$

$$D_{COCA} = -1.486$$

$$D_{CON} = +0.881$$

où R et D sont les composantes du tenseur d'alignement.

## IV-CONSTRAINTES DE LIAISONS HYDROGÈNE

La liaison hydrogène est une interaction électrostatique entre un atome d'hydrogène rattaché à un atome électronégatif et une paire d'électrons libres situés sur un atome (O, S ou N) d'un autre groupement. La distance de la liaison hydrogène varie de 2,7 à 3,1 Å. Les contraintes de ponts hydrogène peuvent être connues en mesurant la vitesse d'échange des protons amides lorsque la protéine est dissoute dans le deutérium. Une série de spectres est alors enregistrée à de faibles intervalles de temps. Le deutérium n'est pas observable dans un spectre RMN  $^1\text{H}$  donc plus le proton est difficilement échangeable avec le solvant plus le temps est long avant la disparition du pic. Nous avons déterminé la vitesse d'échange des pics en mesurant l'évolution de l'intensité des pics au cours du temps. On détermine ainsi la présence potentielle d'un pont hydrogène. Les mesures et analyses sont présentées dans le chapitre « matériels et méthodes » RMN. La Figure 5 indique les liaisons hydrogènes identifiées comme telles et injectées dans le calcul de structure. La contrainte de liaison hydrogène est traitée comme une contrainte de distance dans le calcul mais pour deux distances : la distance O-H et la distance O-N. Ces deux contraintes doivent respecter l'angle formé par la liaison hydrogène.

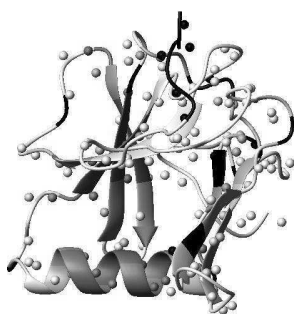


Figure 5 : Visualisation des liaisons hydrogène déterminées à partir du calcul des facteurs de protection [ $\log(k_{\text{ex}}/k_{\text{int}})$ ] décrit dans le chapitre II paragraphe sur la structure RMN de SufA. Plus la couleur tire au violet, plus le  $\text{H}^{\text{N}}$  est protégé.

Ce schéma montre bien que les structures secondaires sont protégées du solvant et qu'elles sont le siège de liaisons hydrogène. De cette analyse « empirique » nous avons construit un fichier de contraintes de liaison hydrogène, H-bond.tbl comprenant 54 contraintes.

## **IV- CALCUL DE STRUCTURE DE SufA**

Les calculs des structures de SufA ont été réalisés suivant un mode de torsion au cours du protocole de recuit simulé du logiciel ARIA version 1.2. La calibration des nOe a été réalisée uniquement à l'aide du système d'auto-calibration contenu dans Aria. L'intensité du pic est convertie en une distance à laquelle est associée une erreur. Le protocole de calcul est présenté dans le chapitre « «matériels et méthodes ».

Les contraintes TALOS sont définies sous forme de contraintes angulaires définies à partir des déplacements chimiques des atomes alors que les contraintes de pont hydrogène ont été définies sous forme de contrainte de type nOe avec une distance de 2.1-3.2 Å. Ces dernières ont été introduites dans les derniers calculs de structures.

Les contraintes RDC sont définies comme telles et on leur associe les composantes du tenseur d'alignement. Ces contraintes ne sont utilisées que dans les deux dernières itérations des calculs.

### **IV.1-Calcul avec ARIA/CNS**

Les contraintes structurales obtenues à partir des données expérimentales ont permis de calculer la structure tridimensionnelle du monomère de SufA. Une vingtaine de calculs ont été nécessaires pour obtenir des structures bien définies.

Tous les calculs de structure ont été réalisés en utilisant systématiquement les contraintes nOe obtenus à partir des attributions de l'ensemble des spectres NOESY-HSQC, ainsi que les contraintes angulaires du fichier TALOS. Les contraintes de liaisons hydrogène et les contraintes RDC n'ont été utilisées que sur les dernières itérations des calculs. A chaque calcul on effectue une analyse des données de sortie tels que les fichiers de violations. On s'intéresse particulièrement, entre autre, aux violations des contraintes non ambiguës en terme de valeur et de fréquence. L'ensemble des violations (des nOe, des angles dièdraux, des H-bonds ou des RDC) aide à trouver les éventuelles erreurs d'attributions. Ces violations nous indiquent les contraintes qui peuvent forcer le calcul à converger vers une structure erronée. Ceci est principalement dû à des erreurs d'attribution, que l'on peut dès lors corriger. A ce stade, on peut se représenter le couple utilisateur/logiciel comme un système expert d'analyse de données. Il est intéressant de suivre l'évolution des violations au cours des différents calculs ainsi que le nombre de nOe attribués et non ambiguës.

Si on prend les calculs décrits dans le paragraphe I.3 ci-dessus on observe que pour la structure intermédiaire on a 5 violations  $> 0.3\text{\AA}$  pour 2272 nOe alors que pour la structure finale, on a 5 violations  $> 0.3\text{\AA}$  pour 2510 nOe.

### **-Structure finale de SufA**

Le dernier calcul de structures a été basé sur 2510 contraintes nOe, 2x116 contraintes d'angles ( $\varphi$  et  $\psi$ ) dièdres obtenus avec TALOS, 54 contraintes de liaison hydrogène et 294 contraintes orientationnelles.

L'analyse du calcul donne donc 2510 nOe non ambiguës dont 807 dans le spectre  $^{15}\text{N}$ -NOESY-HSQC et 1573 dans le spectre  $^{13}\text{C}$ -NOESY-HSQC. Des 2510 nOe, 1191 nOe sont intra-résiduels, 626 séquentiels, 208 nOe entre les protons situés à une distance moyenne et 475 sont à longue distance. 98 nOe sont classés comme ambiguës.

Pour ce dernier calcul de structures, on est parti de la structure étendue de SufA et avons effectué 9 itérations comprenant chacune 200 structures. Les contraintes H-bond sont introduites dans le calcul à la troisième itération et les contraintes RDC à la sixième itération.

L'affinement a été effectué dans l'eau sur les 50 structures de plus basse énergie de la dernière itération. Ne sont conservé que les 20 structures de plus basse énergie à la fin du calcul.

*-Analyse* : Les 20 meilleures structures en terme d'énergie totale de la dernière itération de l'affinement de structure forment l'ensemble RMN final de la structure tridimensionnelle du monomère de la protéine SufA apo. Le RMSD (Root Mean Square Deviation), sur les atomes lourds (N, C $\alpha$  et CO) de la chaîne principale et sur tous les atomes lourds de SufA du résidu 17 au résidu 107, par rapport à la structure moyenne a une valeur égale à 1.19 Å. L'énergie expérimentale totale est égale à -2051.22 kcal.mol<sup>-1</sup>. La contribution des nOe à l'énergie expérimentale est de 341.207 kcal.mol<sup>-1</sup>, celle des H-bonds de 55.6472 kcal.mol<sup>-1</sup>, celle des angles dièdres de 26.6622 kcal.mol<sup>-1</sup> et celle des RDC de 1591.06 kcal.mol<sup>-1</sup>.

Afin de vérifier la qualité stéréochimique des structures de SufA satisfaisant les contraintes structurales, le logiciel PROCHECK (<sup>145</sup>) a été utilisé pour évaluer la géométrie des résidus par rapport aux paramètres stéréochimiques dérivés de structures parfaitement connues d'une banque de données. Les 20 structures de SufA affinées dans l'eau ont été analysées sans les parties N-terminale et C-terminale.

L'analyse nous donne :

82.9 % de résidus dans les régions les plus favorables

12 % de résidus dans les régions autorisées

3.6 % de résidus dans les régions généreusement autorisées.

1.5 % de résidus dans les régions interdites

ces statistiques sont correctes sauf pour le dernier pourcentage. On peut l'expliquer par la flexibilité de la boucle autour de la cystéine 50 et par le manque d'attribution des résidus (84, 85 et 86) de la petite hélice  $\alpha$  (qu'on ne voit pas systématiquement dans nos structures). Ce manque d'information structurale ne permet pas au calcul de converger correctement dans cette zone.

Les 20 structures de plus basses énergies dans l'eau ont été analysées avec le logiciel MOLMOL en ne considérant que les parties structurées de la protéine. La superposition des structures a été faite par rapport à la structure de plus basse énergie en ne considérant que les structures secondaires de la chaîne polypeptidique et les atomes lourds.

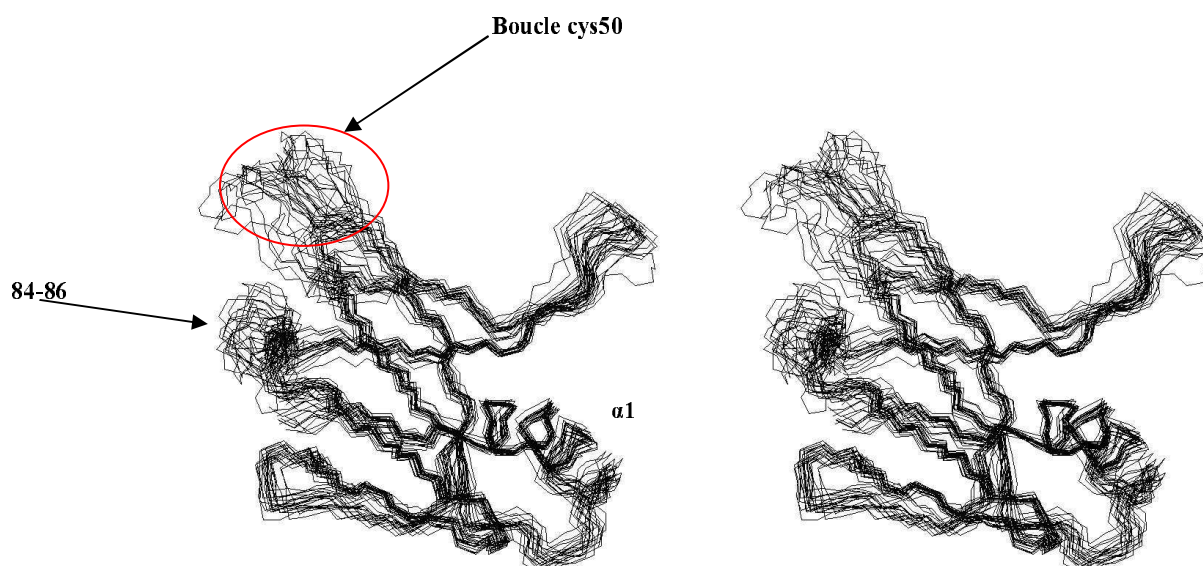


Figure 6 : vue stéréochimique des 20 structures de plus basses énergies de la protéine SufA sans les parties C-terminal et N-terminal après affinement dans l'eau [résidus 17 à 107].

-Analyse du RMSD sur les 20 structures (pour les résidus 17-107)

	<b>RMSD (Å)</b>
Chaîne polypeptidique & structure secondaire & atomes lourds	<b>0.60</b>
Chaîne polypeptidique & structure secondaire & tous les atomes	<b>0.61</b>
Toute la structure & structure secondaire & atomes lourds	<b>1.04</b>
Toute la structure & structure secondaire & tous les atomes	<b>1.27</b>
Chaîne polypeptidique & résidus [17-107] & atomes lourds	<b>1.19</b>
Chaîne polypeptidique & résidus [17-107] & tous les atomes	<b>1.23</b>
Toute la structure & résidus [17-107] & atomes lourds	<b>1.52</b>
Toute la structure & résidus [17-107] & tous les atomes	<b>1.70</b>

Les RMSD présentent des valeurs très acceptables. On trouve bien entendu la valeur de RMSD la plus basse pour l'ensemble des structures secondaires en ne considérant que la chaîne polypeptidique et les atomes lourds.

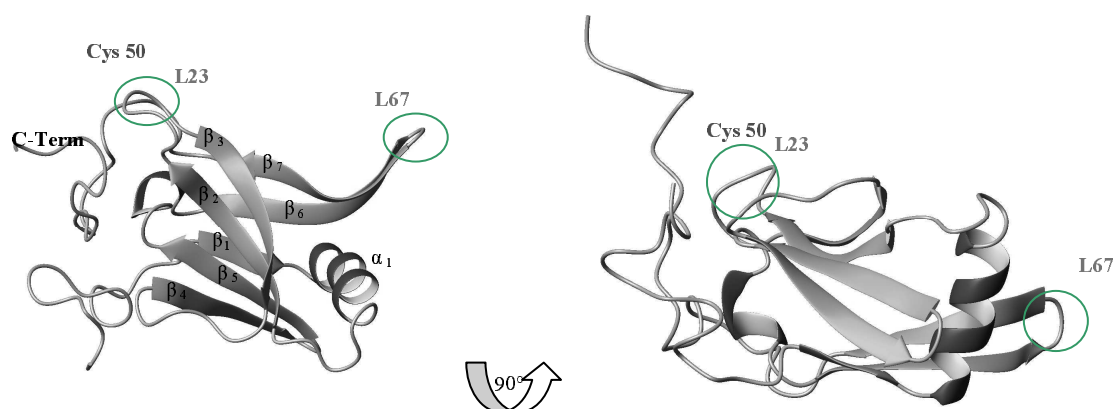


Figure 7 : vue de la structure de plus basse énergie de la protéine SufA sans les parties C-terminal et N-terminal après affinement dans l'eau [résidus 17 à 107] sous deux angles.

La qualité de l'ensemble final de la structure RMN de apo-SufA a été analysée à l'aide du programme PROCHECK-NMR. Pratiquement 96% des angles dièdres  $\Phi$  et  $\psi$  de apo-SufA sont situés dans les régions énergétiquement favorables du diagramme de Ramachandran. L'énergie expérimentale est faible ce qui suggère une bonne coïncidence avec les contraintes expérimentales. Le résumé des statistiques structurales est donné dans le Tableau 2.

#### Contraintes expérimentales RMN

nOe intra-résidus non ambiguës	1191
nOe séquentiel non ambiguës	626
nOe médium non ambiguës	208
Total non ambiguës	2025
Total ambiguës	97
Angles dièdres	2x116
Liaisons hydrogènes	54
Couplages dipolaires résiduels	294

#### r.m.s.d (déviations par rapport aux coordonnées moyennes (Å)<sup>a,b</sup>)

Chaîne polypeptidique & structure secondaire et atomes lourds	0.6
Chaîne polypeptidique & structure secondaire et tous les atomes	0.61
Toute la structure & résidus & atomes lourds	1.52
Toute la structure & résidus & tous les atomes	1.70

#### Violations de contraintes dans plus de 50 % des structures:

Distance (0.3 Å)	10.75
Dièdres (5°)	7.70
RDC (0.5Hz)	15.85
Liaisons hydrogènes	0

**Diagramme du Ramachandran (%)<sup>b,c</sup>**

Résidus dans régions favorables	82.9%
Résidus dans régions autorisées	12%
Résidus dans régions supplémentaires autorisées	3.6%
Résidus dans régions interdites	1.5%

<sup>a</sup> Statistique sur les résidus [17-107]

<sup>b</sup> Moyenne des 20 structures de plus faible énergie

<sup>c</sup> Analyse avec le programme PROCHECK-NMR (les glycines et les prolines ne sont pas prises en considération)

Tableau 2 : Contraintes et statistiques structurales pour le calcul de apo-SufA

Disposant de la structure cristallographique de SufA qui a été publiée entre temps, nous reviendrons sur cet événement, nous avons effectué une superposition des structures de monomère. La superposition a été faite par rapport aux structures secondaires de la protéine SufA sans les parties N-terminal et C-terminal avec l'aide du logiciel MOLMOL et est présentée à la figure 8.



Figure 8: superposition de la structure par RMN de plus basse énergie de SufA en orange avec la structure radiocristallographique (Wada et al (2005) *FEBS letters* [<sup>146</sup>]) en vert (pour les résidus 17-107 de SufA).

On constate aisément que les deux structures monomériques sont similaires. On ne peut rien conclure de probant sur les deux boucles (autour de la cystéine 50 et autour de la glycine 98). En effet nous ne présentons ici qu'une seule structure RMN et nous verrons que



la boucle de la cystéine 50 est flexible alors que celle de la glycine 98 est en interaction avec l'autre dimère. On obtient un RMSD de 1.63 Å (Chaîne polypeptidique & résidus [17-107] & tous les atomes) entre ces deux structures. Si les structures sont proches, on constate toutefois des différences au niveau de la boucle L67 entre les brins  $\beta 6$  et  $\beta 7$  ainsi qu'au niveau de la petite hélice  $\alpha 2$ . Nous savons que la boucle L67 est en fait située à l'interface des deux monomères puisque les résidus qui la constituent présentent des contraintes nOe intermoléculaires. Il se peut dès lors que le dimère présente une structure légèrement différente entre la forme cristallisée et la forme en solution. Il est plus difficile de conclure quant à la présence de la deuxième hélice. En effet, le manque d'attribution des résidus à ce niveau de la structure pénalise le calcul qui n'arrive pas à converger correctement pour cette zone. Il semblerait donc qu'il y ait bien une boucle même si celle présentée sur la structure cristallographique semble atypique. La principale différence entre les deux structures concerne la partie C-terminale où sont localisées les deux cystéines conservées 114 et 116. Wada et *al* observent dans leur structure dimérique qu'un des segments C-terminal vient se positionner à l'interface des monomères (Figure 9). Le même segment de l'autre monomère n'est pas observé par radiocristallographie. Il en est de même pour la boucle de la cystéine 50 conservée qu'ils observent sur le même monomère que le segment C-terminal visible, alors qu'ils ne voient pas cette boucle sur le second monomère. La partie C-terminale et la boucle de la cystéine 50 de l'unité  $\alpha 2$  présentent une densité électronique faible ce qui semble vouloir dire qu'elles sont flexibles tout comme ce qu'on observe par RMN. La flexibilité de la boucle L23 entre les brins  $\beta 2$  et  $\beta 3$ , et des segments N-terminal et C-terminal est confirmée par l'étude de dynamique ci-dessous.

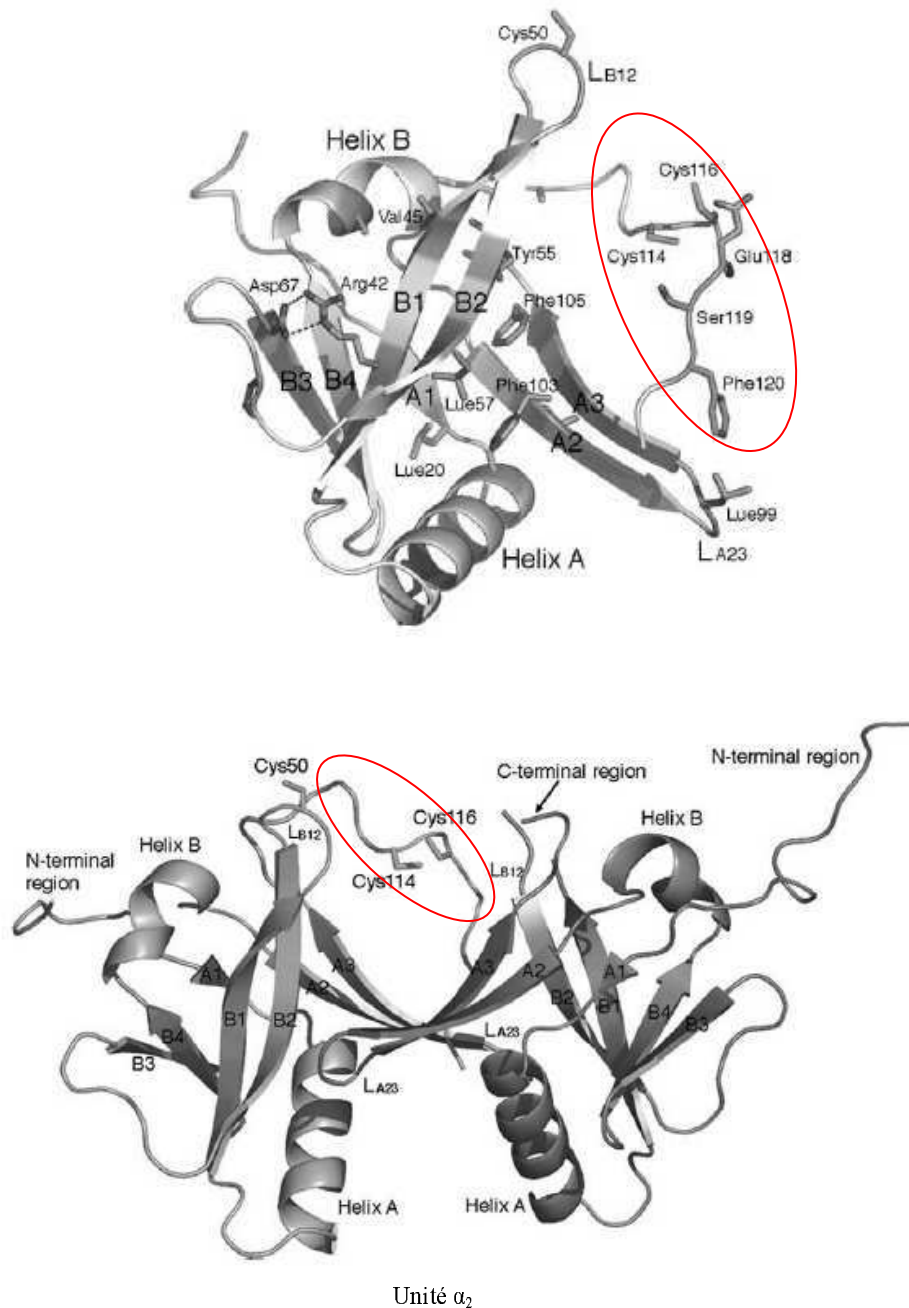


Figure 9 : Structure radiocristallographique de SufA d'*E. coli* (Wada et al (2005) *FEBS letters*)

Encerclé en rouge : segment C-terminal

Si on fait un bilan de ce calcul de structure, on peut dire que nous avons atteint notre objectif en ce qui concerne le monomère. La tâche n'a pas été si aisée du fait que nous ne sommes pas en présence d'un monomère mais d'un dimère. La taille de la protéine devient un handicap qu'il est dès lors nécessaire de contourner. C'est ce que nous avons fait en effectuant le double marquage et en enregistrant des expériences qui pouvaient nous donner des informations complémentaires. Nous avons souffert du manque d'attribution qui a induit un

taux d'ambiguïté élevé au début des calculs ce qui a fortement perturbé ARIA. En effet, après quelques calculs, le calcul de structure ne convergeait pas et il a été nécessaire d'effectuer une grande partie des attributions manuellement, ce qui représente un coût en temps non négligeable. On est peut-être là devant une des limites des algorithmes actuels qui pour le moment n'ont pas acquis l'expertise humaine. Sans ces codes il nous serait toutefois impossible, de résoudre ce type de structure.

D'autre part dans ce genre de calcul, combinant des informations nOe dont la grande difficulté est le calibrage et des informations orientationnelles issues des couplages dipolaires résiduels, il faut jouer sur le poids respectif des deux types de contraintes à introduire. Il faut pour cela être vigilant, au cours des calculs où sont introduits les deux types de contraintes, aux violations respectives de ces contraintes. Trop de poids mis sur une contrainte peut créer des violations systématiques de l'autre et inversement. Ceci dit, les RDC sont des contraintes locales alors que les nOe sont à la fois des contraintes locales et longues portées. Le conflit risque donc plus entre les nOe courte portée et les RDC qu'entre les nOe longue portée et les RDC. En effet une mauvaise calibration de nOe longue portée affectera peu les RDC. Par contre une mauvaise calibration de nOe courte portée affectera nécessairement les RDC. La conclusion est peut être de donner plus de poids aux nOe dans les premières itérations pour inverser la tendance sur les dernières (en jouant sur les constantes de forces respectives de deux contraintes).

### *-Conclusion*

Comme le montre la représentation en ruban de la figure 7 ; la protéine SufA est repliée en un simple domaine contenant une hélice  $\alpha$  formellement identifiée et 7 brins  $\beta$ . Le cœur central de la protéine est constitué de deux feuillets  $\beta$ , un feuillet de 3 brins (B1, B6, B7) et un feuillet de 4 brins (B2, B3, B4, B5) au côté des desquels on trouve l'hélice  $\alpha_1$ . Deux boucles caractéristiques, une, nommée L67, entre les brins B6 et B7 (résidus 99-100) et une, nommée L23, entre les brins B2 et B3 (résidus 49-51). Les parties C-terminal (résidus 108-130) et N-terminal (résidus 1-16) ne sont pas structurées. Les cystéines conservées 114 et 116 dans la partie C-terminale sont parfaitement identifiées mais ne sont pas structurées. La cystéine 50 est située dans la boucle L23 qui présente une flexibilité importante. Les structures caractéristiques B2-L23-B3 et B6-L67-B7 sont stabilisées par des interactions hydrophobes impliquant les résidus Leu20, Val45, Leu57, Phe103 et Phe105 à l'intérieur de la molécule.

## V-ANALYSE DE LA DYNAMIQUE DE SufA

Dans le but de compléter la caractérisation du monomère de SufA apo, nous avons étudié la relaxation hétéronucléaire N-H à 600 MHz. Pour l'étude de la dynamique de SufA, nous avons mesuré les trois paramètres de relaxation qui sont la vitesse de relaxation longitudinale  $R_1$ , la vitesse de relaxation transversale  $R_2$  et les paramètres de relaxation  $\rho$  et  $\sigma$ .

### -Extraction des paramètres de relaxation

Nous avons extrait les valeurs de  $R_1$  et  $R_2$  à partir de la mesure des intensités de pics de corrélations  $^1\text{H}$ - $^{15}\text{N}$  HSQC enregistrées pour différents délais de relaxation. Les paramètres de relaxation ont pu être mesurés pour 85 % des N-H de la chaîne polypeptidique. Les sites qui n'ont pas été considérés sont les prolines ou des résidus dont les pics de corrélation ont des recouvrements spectraux, ce qui induit une imprécision dans la mesure.

La Figure 10 représente le rapport  $R_1/R_2$  ainsi que les temps de relaxation  $T_1$  et  $T_2$ .

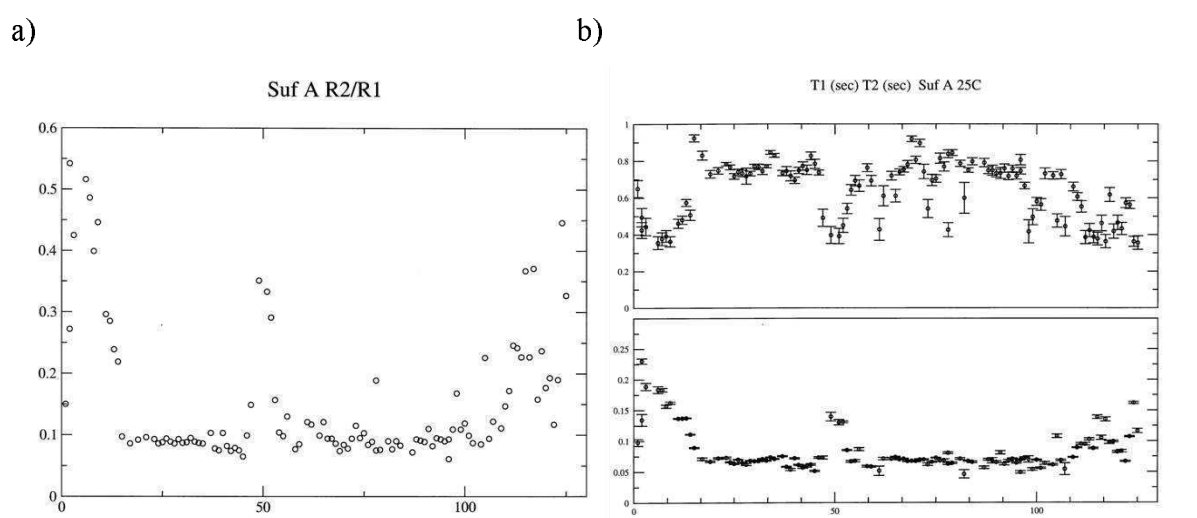


Figure 10 :

- a) L'analyse du rapport  $R_1/R_2$  montre que les parties N-terminal et C-terminal sont flexibles et par conséquent pas structurées. La boucle où se situe la cystéine 50 présente aussi une dynamique importante.
- b) Les mesures des temps de relaxations  $T_1$  et  $T_2$  montrent de même une flexibilité des parties terminales ainsi que de la boucle de la cystéine 50.

*-Commentaire*

Du rapport  $R_1/R_2$  et des mesures de  $T_1$ ,  $T_2$  nous pouvons dire que les segments N-terminal et C-terminal sont flexibles tout comme la cystéine 50 et les résidus qui l'entourent. En effet, ces paramètres intègrent des informations de dynamique qui influencent la relaxation de chacun des noyaux d'un vecteur H-N d'un résidu  $i$ .

A ce stade de la thèse nous n'avons pas approfondi ces mesures de relaxation. Ces mesures ont été faite sur le dimère, il nous paraît donc prioritaire de résoudre la structure dimérique de la protéine SufA pour exploiter au mieux les données de relaxation, d'autant plus que SufA n'a pas une forme réellement globulaire comme semble le montrer les premières déterminations de structure sur le dimère.

De ces figures, nous pouvons dire que les parties N-terminal et C-terminal sont flexibles ainsi que la cystéine 50 et les résidus qui l'entourent.

## Chapitre IV

## Discussion

La protéine SufA présente une structure dimérique d'environ 2x14 kDa avec 130 acides aminés par monomère. La qualité des premiers spectres  $^1\text{H}$ - $^{15}\text{N}$  HSQC obtenus sur SufA apo nous a encouragé à poursuivre cette étude. La stratégie suivie repose sur la préparation d'un échantillon doublement marqué et l'enregistrement de l'ensemble des spectres nécessaires à l'attribution de la protéine. La première étape d'attribution de la chaîne polypeptidique a donné de bons résultats, alors que ceux de l'étape d'attribution des chaînes latérale (64.5% pour le  $^1\text{H}$ ) sont à la limite de ce qui est acceptable en terme d'attribution (score > 75% pour le  $^1\text{H}$ ). Néanmoins ils sont de niveaux suffisants pour envisager un calcul de structure. Cette perte d'information a été un handicap important, particulièrement pour les premiers calculs de structures basses résolutions. En effet, nous souhaitions utiliser toute la potentialité du logiciel ARIA, capable d'effectuer une auto-attribution des contacts nOe, pour réaliser les calculs. Mais ces calculs initiaux n'ont pas permis de converger vers une structure stable. L'importance du nombre d'ambiguïté généré par les atomes non attribués, tout comme les erreurs éventuelles de l'attribution manuelle, n'ont pas permis aux différents calculs de converger. Il a donc été nécessaire d'attribuer manuellement les contacts nOe pour permettre au calcul de replier la structure de la protéine. Bien entendu, ce type d'attribution doit être effectué sans générer d'erreurs (qui risquent alors de fausser le calcul). C'est, en quelque sorte, la stratégie suivie classiquement pour la détermination de structures de macromolécules : attribution de contacts nOe, dont on est sûr, suivie d'un calcul préliminaire

(à partir de ces contraintes) avec le programme CNS d'une structure basse résolution, qui servira de modèle dans la suite du calcul sous ARIA. On peut cependant se demander pourquoi, lorsque nous avons utilisé comme modèle la structure de la protéine homologue IscA, le calcul ne convergeait pas mieux. En effet, par ce biais, nous suivions une stratégie classique. Il semble évident que les erreurs d'attributions manuelles ont grandement pénalisées les calculs initiaux, toutefois elles n'ont pas toujours été rejetées au cours des calculs car elles n'étaient pas systématiquement violées. C'est à ce niveau que la perte d'informations structurales, lié au manque d'attribution, était la plus pénalisante. Il semble légitime de penser qu'avec des contraintes supplémentaires, on aurait pu identifier plus facilement les erreurs d'attribution, ou qu'elles auraient été rejetées.

C'est donc seulement après avoir attribué un maximum de contacts nOe manuellement que les premiers calculs ont abouti à une convergence de structure. Le problème est que cette attribution manuelle risque d'introduire de nouvelles erreurs qui peuvent contraindre la structure à se replier dans une mauvaise conformation. Toutefois, les contacts nOe ne sont pas les seules contraintes introduites dans le calcul et par conséquent on devrait observer des violations systématiques de ces autres contraintes lorsque les nOe orientent le calcul vers un mauvais repliement. En minimisant les erreurs, les algorithmes de calcul comme ARIA apportent une aide très précieuse pour les identifier de manière à, soit de les éliminer, soit de les réattribuer.

La détermination de structure est, à l'heure actuelle, un travail en binôme entre la machine et l'utilisateur. Ce dernier garde, bien entendu, la maîtrise d'œuvre de l'étude mais ne pourrait pas résoudre de structure de macromolécule sans algorithme et moyens informatiques puissants. Et ceux-ci progressent régulièrement avec le développement de nouveaux moyens d'expertise de spectres nOe. Ainsi, Hermann et *al* ont développé un nouveau logiciel, nommé ATNOS-CANDID [<sup>147</sup>], capable de gérer les spectres RMN NOESY-HSQC bruts qui ne sont absolument pas attribués. Ce logiciel effectue automatiquement le « peak picking » et l'attribution des contacts nOe suivant une procédure de validation de l'attribution des contacts nOe plus efficace que celle d'ARIA, via un processus nommé « Network Anchoring ».

Cette procédure reste toutefois réservée à des protéines pour lesquelles les spectres sont relativement bien résolus avec un bon rapport signal sur bruit. D'autre part, il est toujours nécessaire d'effectuer l'attribution des atomes de la protéine. Or dans le cas de SufA, c'est à ce niveau que nous avons rencontré le plus de problèmes. Comment ce comporterait ce type

d'algorithme sur une protéine comme SufA ? Il serait intéressant de le tester d'autant plus que SufA est un dimère.

En ce qui nous concerne, nous poursuivons le calcul de structure de la protéine SufA sous sa forme dimérique avec ARIA1.2 à partir des contacts intermoléculaires identifiés comme sûrs et d'autres contacts nOe possédant une ambiguïté, intra ou inter moléculaire. Nous avons ensuite fait le choix de partir de la structure monomérique de plus basse énergie comme structure modèle, en adaptant sa topologie au calcul du dimère bien entendu. Surtout nous avons fixé les contraintes de distance de ce monomère modèle. La seule possibilité laissée au calcul est de positionner les deux monomères à partir des contraintes intermoléculaires identifiées comme telles et de gérer les contraintes ambiguës. Bien entendu, toutes les autres contraintes (angles dièdres, liaisons hydrogène, RDC) sont introduites dans le calcul sous ARIA. Nous ne développerons pas plus cette partie à ce stade de la thèse puisque nous venons de commencer les calculs. Une autre stratégie envisageable est d'effectuer une analyse par « docking ». En effet le nombre de contacts nOe intermoléculaires non ambiguës pour la protéine SufA est faible et ils sont localisés uniquement entre les résidus 98-99 d'un monomère (boucle L98) et les résidus 54-55 du brin  $\beta$ 3 du second monomère (cf Figure 7 chapitre IV). D'autre part les contraintes ambiguës sont situées au niveau d'un segment 113-116 de la partie C-terminale des deux monomères. Ces contraintes sont toutefois peu nombreuses. L'information structurale est donc faible et par conséquent, le calcul risque de ne pas aboutir au positionnement des monomères. L'utilisation d'un logiciel de « docking » auquel on associe les contraintes nOe expérimentales que nous avons collectées devrait permettre de trouver la solution plus rapidement. En effet il ne s'agit plus de ne recourir qu'aux seules données expérimentales, mais des les introduire comme contraintes ambiguës pilotant le processus de « docking ». Ce processus, utilisé notamment pour la détermination de complexes protéiques par exemple, est basé sur des calculs *ab initio*. La plupart de ces programmes utilisent la même approche : une molécule est fixée dans l'espace et la seconde est déplacée autour de la première. A chaque nouvelle configuration, un « score » est calculé sur la base de termes tels que l'interaction électrostatique, la répulsion de van der Waals. Mais le calcul doit explorer l'ensemble de l'espace conformationnel disponible. L'idée des créateurs du logiciel « HADDOCK » (High Ambiguity Driven protein-protein DOCKing) [148] est de mélanger les deux processus. Ainsi les contraintes expérimentales orientent le calcul *ab initio*. Cette stratégie devrait, dans notre cas, aboutir plus rapidement à la conformation du dimère.



La détermination de la structure de la protéine sous sa forme dimérique pourrait être une étape essentielle pour comprendre comment est chélaté le centre [Fe-S] au sein de la protéine SufA. En effet il n'a jamais été possible, jusqu'à présent, de déterminer précisément et avec certitude les ligands du centre [Fe-S] que ce soit par des études de mutagenèse dirigée (cf § III.3.3 chapitre I) ou à partir des structures cristallographiques des protéines apo IscA et apo SufA d'*E. coli* [132,146]. Toutefois, l'obtention récente de la structure de la protéine homologue IscA de *Thermosynechococcus elongatus* [<sup>149</sup>] sous une forme holo a permis de définir ces ligands.

Il existe deux structures de IscA apo d'*E. coli* qui présentent une protéine sous forme de dimère de dimères : dans une des structures, Bilder et al [<sup>150</sup>] décrivent deux types d'assemblage de dimères (Figure 1a) alors que dans la seconde structure, Cupp-Vickery et al [132] décrivent un seul type d'assemblage (Figure 1b).

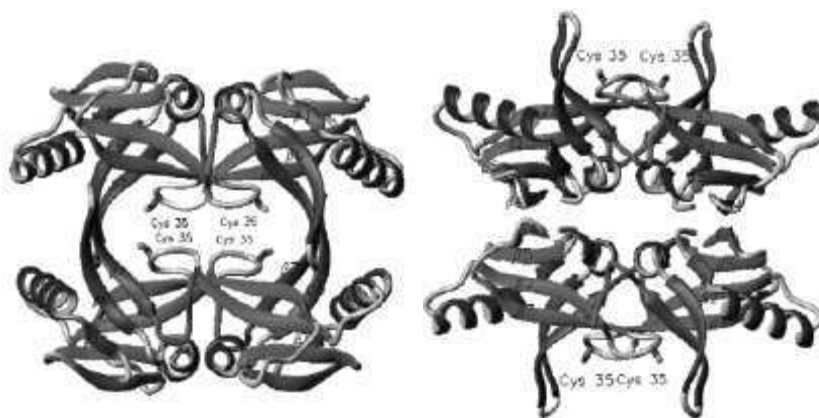


Figure 1a : Structure cristallographique de IscA d'*E. coli* (Bilder et al (2004) *Biochemistry*)[150]  
Deux types de tétramère : nommé A à gauche et B à droite

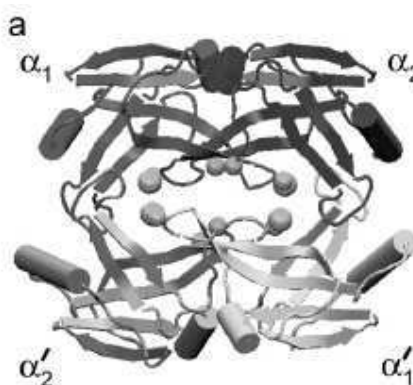


Figure 1b : Structure cristallographique de IscA d'*E. coli* (Cupp-Vickery et al (2004) *Journal of molecular biology*) [132]

Un seul type de tétramère – en vert Cys35, en jaune Glu98.

Bider et *al* proposent que le tétramère A soit la forme physiologique présente en solution. Ceci semble confirmé par le second groupe, qui n'observe qu'un tétramère de structure proche de celui du type A de Bilder et *al*.

Quoiqu'il en soit aucune de ces structures n'a permis d'observer la partie C-terminale où sont positionnées deux des trois cystéines conservées, les cystéines 99 et 101 (équivalentes à 114 et 116 dans SufA) supposées coordonner le centre [Fe-S].

La structure cristallographique de la SufA sous sa forme apo a été obtenue avec une résolution de 2.7 Å par Wada et *al* (Figure 3). Cette structure présente, contrairement à la protéine homologue IscA une forme dimérique. La structure monomérique de cette protéine a été présentée dans le chapitre précédent mais nous pouvons rappeler ici que le dimère est constitué de deux monomères positionnés symétriquement. Pour un des monomères, le segment C-terminal comprenant deux des cystéines conservées (cys114 et cys 116) est positionné dans la zone d'interaction des deux monomères. La boucle comprenant la cystéine 50 conservée est observable sur le même monomère. Pour le second monomère aucune de ces régions n'est observée.

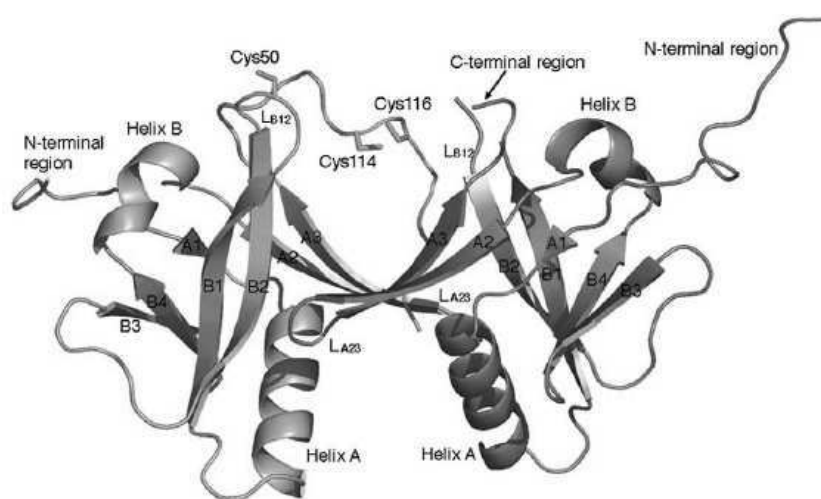


Figure 3 : Structure cristallographique de SufA d'*E. coli* (Wada et *al* (2005) *FEBS letters*)

Comme mentionné précédemment la structure de SufA existe, contrairement à IscA, sous forme de dimère. Il est important de noter que s'il n'existe pas de différences fondamentales entre les monomères des deux protéines IscA et SufA, la situation est différente dans le cas des dimères. En effet, les zones d'interactions des dimères pour les deux

protéines sont complètement différentes. Si on compare par exemple le dimère  $\alpha_1\alpha_2$  de la structure de Cupp-Vickery (Figure 4B) à la structure de SufA obtenue par Wada et al (Figure 3), on constate que la zone d'interaction qui implique les résidus du brin A3 et du segment C-terminal ainsi que les contacts entre la boucle L<sub>23</sub> et les brins B1, B2 et A3, sont inexistantes dans la structure de Cupp-Vickery. Le dimère de SufA présenterait toutefois une ressemblance avec l'unité dimérique  $\alpha_1\alpha_2$  déduite de la forme tétramérique d'IscA obtenue par Vickery et al (Figure 4D). Mais ce n'est pas la forme retenue pour former le centre [Fe-S] suite à la modélisation. Le dimère C correspond à la seule forme où les sous unités sont assez proches pour former le centre [Fe-S] comme modélisé par Vickery et al. Mais ce positionnement ne peut exister que dans l'état tétramérique et il ne présente aucune ressemblance avec le dimère de SufA.

On est donc en présence de deux systèmes biologiques proches mais qui ont des états oligomériques différents.

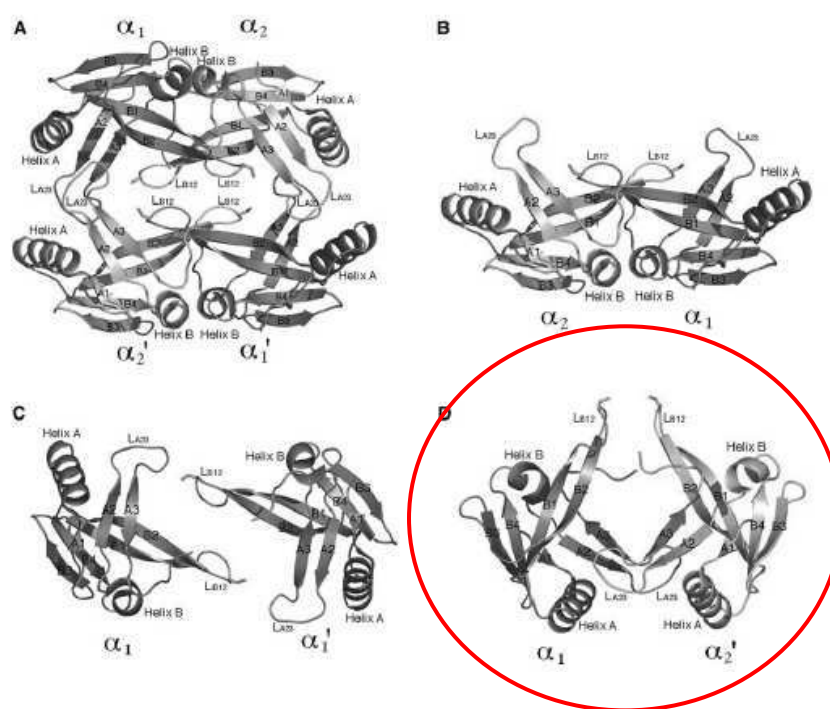


Figure 4 :

A tétramère d'IscA d'*E. coli* obtenu par Cupp-Vickery [132].

B, C, D dimères déduits de la structure tétramérique d'IscA d'*E. coli* (Wada et al (2005) *FEBS letters*) [146]

Dernièrement, la structure de la protéine IscA holo de *Thermosynechococcus elongatus*, obtenue sous forme native avec son centre [Fe-S], a été résolue par radiocristallographie par Morimoto et *al* [149], avec une résolution de 2.5 Å. Cette structure présente un ensemble formé de deux unités dimériques, chacune contenant un centre [2Fe-2S] positionné à l'interface (Figure 5). Dans cette configuration, le centre [2Fe-2S] est chélaté par la cystéine 103 de la partie C-terminale (cystéine 116 pour SufA) d'un monomère et par les trois cystéines conservées (Cys 35, Cys 101, Cys 103) d'un deuxième monomère mais positionné de manière dissymétrique par rapport au premier. A partir de cette structure et sur la base d'expériences biochimiques de la protéine en solution, Morimoto et *al* proposent que seule la structure dimérique dissymétrique  $\alpha\beta_{sw}$  avec un centre [2Fe-S] à l'interface, soit la forme active en solution (Figure 6). Toutefois il n'exclue pas le fait que l'autre forme dissymétrique  $\alpha\beta'_{sw}$  avec un centre [Fe-S] puisse également exister en solution.

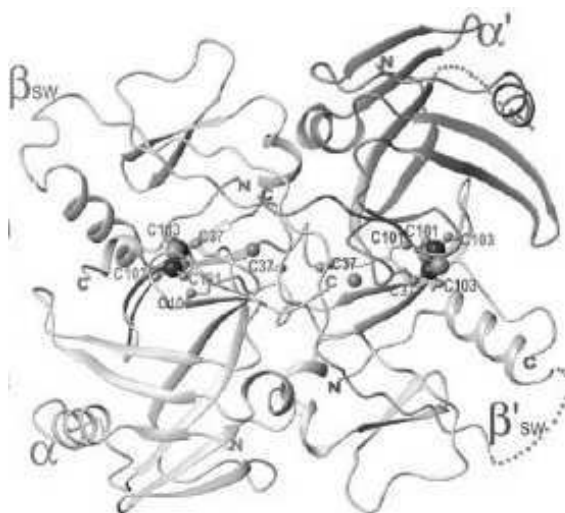


Figure 5 :Structure cristallographique de IscA de *Thermosynechococcus elongatus* (Morimoto et *al.*, (2006) *Journal of molecular biology*) [149].

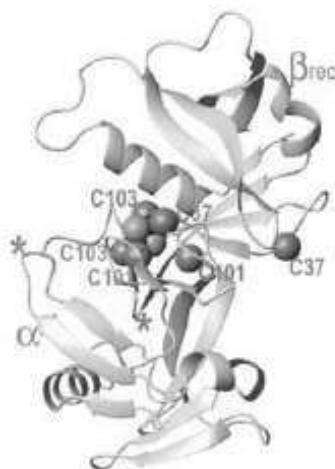


Figure 6 :Modèle proposé de IscA de *Thermosynechococcus elongatus* chélatant un centre [2Fe-2S] (Morimoto et *al.*, (2006) *Journal of molecular biology*) [149].

Sur la base de sa structure, Cupp-Vickery et *al* proposent quant à eux un modèle de la protéine sous une forme holo avec un centre  $[2\text{Fe-2S}]$  à l'interface de deux monomères, chaque fer étant coordonné par les deux cystéines C-terminal de chaque chaîne polypeptidique. Dans ce modèle la cystéine conservée, en position 35, (équivalent à la cystéine 50 d'*E. coli*) n'est pas impliquée (Figure 2). Il est à noter que leur proposition de structure avec le centre  $[\text{Fe-S}]$  exclue toute possibilité de formation du centre  $[\text{Fe-S}]$  à l'interface d'un tétramère.

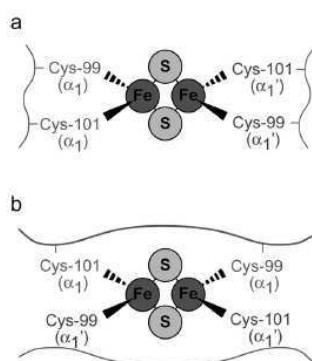


Figure2 : Modèles calculés par Cupp-vickery pour la chélation d'un centre  $[2\text{Fe-2S}]$

D'autre part, sur la base de la structure cristallographique de SufA, Wada et *al* ont modélisé le site intermoléculaire où le centre  $[\text{Fe-S}]$  pourrait être positionner. Pour ce faire, ils ont d'abord modélisé le segment C-terminal manquant, puis en tenant compte de contraintes géométriques (distances Fe-S) intinsèques aux centres  $[2\text{Fe-2S}]$  et  $[4\text{Fe-4S}]$ , ils ont analysé les distances possibles entre le centre métallique et les résidus potentiellement ligand. Ainsi, ils ont déterminés que les Cys 114 et 116 de chaque segment C-terminal participent à la chélation du centre  $[\text{Fe-S}]$ . Par contre, ils ne donnent aucune information sur la participation de la troisième cystéine conservée, la Cys 50. Or, nous avons vu, que dans la protéine holo SufA reconstituée, les trois cystéines sont proches du centre métallique. Nous avons donc là une information qui correspondrait plus à celle obtenue par Morimoto et *al*, mais d'un autre côté notre structure semble très proche de celle de Wada et *al*.

Tous ces résultats sont, bien entendu, à prendre en considération pour qui s'intéresse à ces systèmes et d'autant plus que les protéines IscA et SufA présentent une grande homologie de séquence et de structures secondaires. Mais ces résultats de cristallographie montrent qu'il

existe des différences importantes entre l'état oligomérique de la protéine IscA cristallisée et celui en solution. D'autre part, ces structures de centres [Fe-S] s'appuient toujours sur des modèles calculés ou déduits à l'exception de la structure de *T. elongatus*. Au final, nous n'avons toujours pas accès à une vision limpide de structures de protéines d'échaffaudage avec son centre [Fe-S]. L'obtention d'une deuxième structure avec un centre [Fe-S] semble donc importante pour valider ou infirmer les résultats obtenus par Marimoto *et al.*

L'hypothèse forte qui se dégage, globalement, de ces études est qu'une entité dimérique avec un centre [Fe-S] à l'interface représente la forme holo des protéines de type A

La résolution de la structure par RMN sera donc une aide précieuse pour identifier le site de chélation du centre [Fe-S] de SufA. En effet, on pourrait envisager d'effectuer, par exemple, des expériences de RMN paramagnétique sur la protéine holo SufA.

La détermination de la structure de SufA en solution est donc l'étape préalable à cette investigation future. Reste à déterminer la structure de la protéine dimérique. Et ce même s'il existe déjà la structure radiocristallographique de la protéine SufA d'*E. coli* [132]. En effet, comme nous l'avons indiqué sous sa forme apo, l'ensemble de la protéine pourrait être structuré (du moins sous sa forme cristallisée). Mais ce n'est pas ce que l'on observe par RMN avec SufA. En effet, les résultats de dynamique montrent au contraire que les parties N-terminal et C-terminal sont flexibles, tout comme la boucle de la cystéine 50. Par contre, nous observons la même zone d'interaction du dimère. En effet, des contacts nOe intermoléculaires entre le résidu 98 d'un monomère avec le résidu 58 de l'autre monomère de la protéine SufA ont été identifiés avec certitudes. D'autres contacts nOe sembleraient conforter cette zone d'interaction mais il existe encore des ambiguïtés entre intra et inter moléculaire pour ces derniers. Le calcul du dimère prend donc ici toute sa valeur.

## Conclusion

On voit bien ici toute la complémentarité que peuvent avoir la RMN et la radiocristallographie. En effet, l'une et l'autre sont susceptibles d'apporter un certain nombre d'informations structurales qui peuvent se confirmer ou se compléter. La RMN donne accès à l'étude de la dynamique de la macromolécule et peut permettre l'étude d'interaction avec des partenaires physiologiques. C'est une technique qui permet l'étude de la protéine d'intérêt en solution donc dans des conditions proches du milieu physiologique.

En ce qui concerne SufA, l'obtention d'une structure ouvre la porte à d'éventuelles études telles que l'identification des ligands du centre [Fe-S]. On peut aussi envisager des études d'interactions de la protéine avec une protéine cible. Cependant ceci nécessite de bénéficier d'un échantillon de protéine reconstituée homogène au niveau de son centre [Fe-S], ce qui est une des principales difficultés de ces systèmes qui par essence sont flexibles et capables de fabriquer différents types de centres [Fe-S] comme décrit dans le chapitre I. Pour conclure nous présentons notre protéine en espérant pouvoir compléter les informations manquantes...



Structure d'un monomère de SufA obtenue par RMN

## Seconde partie

### La SOR de *Treponema pallidum*

Le principal objectif de l'étude par RMN de la Superoxyde Réductase (SOR) est la caractérisation structurale de son site actif. En effet, outre l'intérêt que représente la détermination de la structure de ce type de système paramagnétique, on espère aider à la compréhension du mécanisme catalytique par l'analyse RMN de mutants dirigés de la SOR, portant sur des résidus du site actif. D'autre part, ce mécanisme fait intervenir un processus de protonation pour lequel le donneur de proton n'est pas connu. Ainsi des expériences RMN de dépendance en pH devrait nous permettre d'identifier le résidu donneur de proton. Le préalable à ces différentes études étant toutefois d'obtenir la structure de la protéine sauvage. Ce travail a été effectué dans un premier temps en collaboration avec le Docteur Jacques Gaillard du laboratoire SCIB au CEA Grenoble. Cependant, des difficultés, principalement liées à la taille de la protéine et à son état paramagnétique, auxquelles s'ajoute la présence de plusieurs conformations de la protéine, ne nous ont pas permis de mener ce projet à terme.

## I-CONTEXTE BIOLOGIQUE

Le radical anion superoxyde est le produit toxique de la réduction monoélectronique de l'oxygène moléculaire. Depuis des années, le seul système connu d'élimination du superoxyde était la superoxyde dismutase, qui catalyse la dismutation du radical anion superoxyde en peroxyde d'hydrogène et en dioxygène [<sup>151</sup>]. Toutefois, il a été récemment découvert que  $O_2^{\cdot -}$  pouvait être aussi éliminé par réduction en peroxyde d'hydrogène, réaction catalysée par la SOR, présente dans certains micro-organismes anaérobies et microaérophiles [<sup>152</sup>, <sup>153</sup>, <sup>154</sup>, <sup>155</sup>, <sup>156</sup>]. Jusqu'à présent, trois classes de SOR ont été décrites.



SOR de classe I : les SORs de *Desulfovibrio desulfuricans* et *Desulfoarculus baarsii* sont les prototypes des SOR de classe I [152]. Leur structure tridimensionnelle (Figure 1) a été déterminée par cristallographie avec une résolution de 1.9 Å pour la SOR de *D. desulfuricans* [156] et de 1.15 Å pour la SOR de *D. baarsii* [157]. Cette molécule est un homodimère de masse moléculaire de 2 x 14 kDa. Le monomère est organisé en deux domaines, appelés centre I et centre II, chacun avec un centre à fer mononucléaire. Le centre I contient un fer ferrique de type rubrédoxine distordu coordonné par quatre cystéines. Le centre II possède un fer ferreux avec une coordination pyramidale à base carrée, où quatre azotes d'histidines sont les ligands équatoriaux et le soufre d'une cystéine, le ligand axial [156]. Les études de réactivité avec  $O_2^{\bullet-}$  ( $O_2^{\bullet-} + 1 e^- + 2H^+ \rightarrow H_2O_2$ ) ont montré que ce centre atypique  $[Fe^{II}(N_{His})_4(S_{Cys})]$  est le site actif de la superoxyde réductase [152].

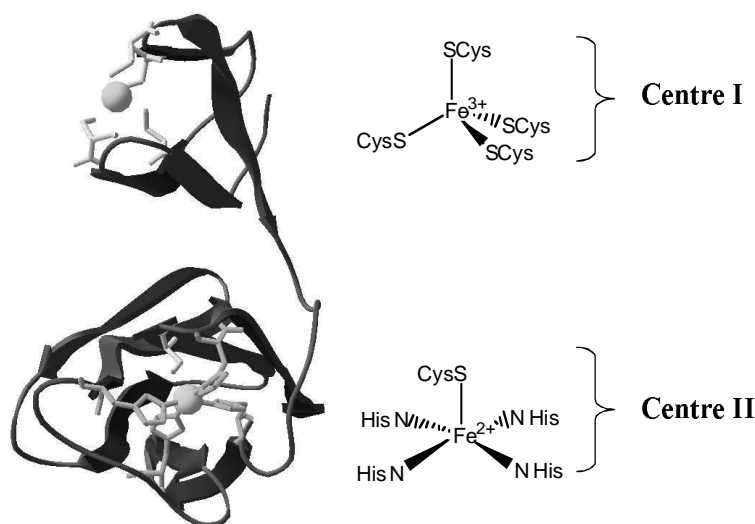


Figure 1 : Structure tridimensionnelle d'une unité monomérique de la SOR de *D. desulfuricans* (Coelho *et al.*, 1997)

SOR de classe II : Ces enzymes isolés chez *D. gigas* et chez *Pyrococcus furiosus* [158], comme la SOR de *T. pallidum*, contiennent un centre  $[Fe^{II}(N_{His})_4(S_{Cys})]$ . Par contre, elles ne possèdent pas de centre de type rubrédoxine  $[Fe^{II}(S_{cys})_4]$  du fait qu'il leur manque le domaine protéique correspondant à la partie N-terminal des SORs de bactéries sulfato-réductrices.

SOR de classe III : L'enzyme de *Treponema pallidum* (2 x 14 kDa) représente les SOR de classe II ; elle ne possède qu'un seul centre à fer localisé dans le domaine C-terminal homologue du centre II des classes I [152,155].

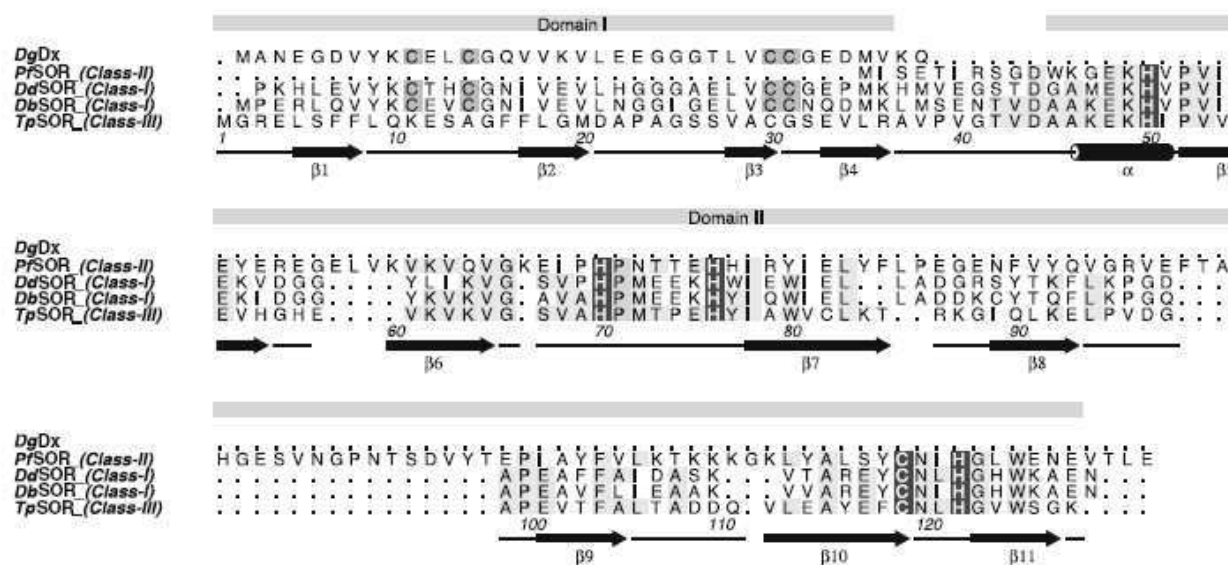


Figure 2 : Alignement de séquence des différentes classes de SOR [159]

On observe sur l'alignement de séquences (Figure 2), que celle de *T. pallidum* peut être alignée avec les séquences de toutes les SORs de bactéries sulfatoréductrices, incluant ce domaine N-terminal. D'autre part chez les SORs de classes deux il existe une boucle supplémentaire de 20 acides aminés dans le domaine C-terminal, qui n'est pas présente chez les autres SORs. Enfin, les SORs de classe I sont des dimères alors que celles de classe deux sont des tétramères [159]. Quant à la SOR de *T. pallidum* c'est un dimère. A cause de ces différences entre les différentes classes de SOR, on pouvait légitimement se demander si, en conséquence, il existait des différences structurales. Il semblait donc pertinent de vouloir déterminer la structure de la protéine SOR de *T. pallidum*, structure qui par ailleurs pouvait apporter une aide à la compréhension du mécanisme réactionnel.

En ce qui concerne ce mécanisme réactionnel (Figure 3), des études de radiolyse pulsée sur la SOR de *D. baarsii* ont apporté des informations déterminantes sur le fonctionnement des ces enzymes, [160, 161, 162, 163, 164, 165]. Il est maintenant généralement admis que la réaction de  $O_2^-$  avec le fer ferreux du centre II se fait via un mécanisme sphère interne. La première étape de la réaction consiste en une réaction bi-moléculaire de la SOR avec  $O_2^-$  au cours d'un processus très rapide contrôlé par la diffusion ( $10^9 \text{ M}^{-1} \text{ s}^{-1}$ ), pour former un premier intermédiaire. Celui-ci est proposé comme étant un fer-peroxo ( $Fe^{3+}-OO^-$ ). L'étape suivante a clairement été identifiée comme un processus de protonation, permettant la

formation d'un second intermédiaire réactionnel [164,165], un fer-hydroperoxo ( $\text{Fe}^{3+}\text{-OOH}$ ), avant la formation des produits finaux de réaction,  $\text{H}_2\text{O}_2$  et le fer oxydé du centre II. Cependant, la cinétique de cette dernière étape du mécanisme réactionnel n'est toujours pas connue [161, 162, 165]. Par ailleurs d'autres groupes proposent que le premier intermédiaire conduit directement à la formation des produits finaux sans passer par la formation d'un second intermédiaire [160, 164].

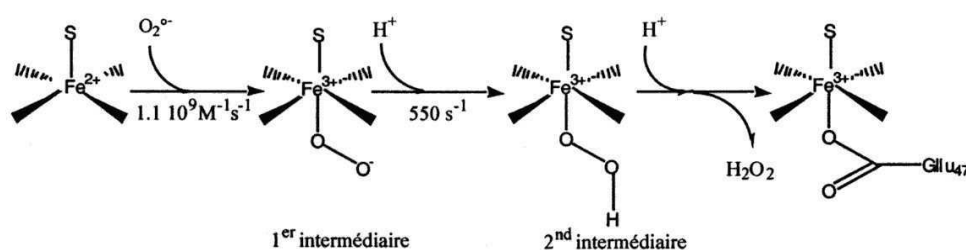


Figure 3 : Mécanisme réactionnel de la SOR

Quand nous avons commencé le projet, aucune structure tridimensionnelle (3D) de la SOR de *T. pallidum* n'était connue. D'autre part les différences significatives entre les deux protéines SOR de *T. pallidum* et de *D. desulfuricans*, nous ont décidé à déterminer cette structure par spectroscopie RMN. Comme nous l'avons noté précédemment il n'y a pas de centre I (type rubrédoxine et de fonction inconnue) dans la SOR de *T. pallidum*, par contre elle possède une extension N-terminale qui n'existe pas chez la SOR de *D. desulfuricans*. Rappelons l'existence d'une troisième classe de SOR qui ne possèdent pas non plus l'extension N-terminale mais qui forment des tétramères [158]. Ces variations dans les structures des différentes SOR posent quelques questions. Quel est le rôle de ce centre I, comme celui de l'extension N-terminale pour la SOR de *T. pallidum* ; est-elle structurée et cela influence-t-il la fonction de la protéine ?

## II-MATÉRIELS ET MÉTHODES

Le protocole de production de la protéine SOR est décrit dans le chapitre matériels et méthodes où on trouvera par ailleurs les caractéristiques physico-chimiques de la SOR de *T. pallidum*.

A l'issue de la purification, la protéine est dans un tampon Tris-HCl de pH égal à 7.6. L'analyse par RMN est effectuée dans de l'eau lourde ( $D_2O$ ). Pour ce faire, le solvant initial a été échangé par lavages successifs de l'échantillon de protéine avec une solution de  $D_2O$  tamponnée à  $pD = 6.8$  par un tampon phosphate à 50 mM. A ce stade, l'échantillon RMN contient environ 1 mM de protéine et présente une coloration bleue très caractéristique du centre métallique du site actif dans un état oxydé  $Fe^{3+}$ . Pour obtenir la forme active de la protéine,  $Fe^{2+}$ , qui nous intéresse, la réduction a lieu dans des conditions d'anaérobiose en boîte à gants. L'échantillon réduit est ensuite transféré dans un tube RMN haute précision qui est fermé par un système étanche de type valve Young<sup>TM</sup>.

Les expériences RMN ont été effectuées sur un spectromètre Varian Unity Plus opérant à 500 MHz et équipé d'une sonde de détection dite inverse [166].

Les déplacements chimiques RMN du proton ont été référencés relativement au signal de résonance de l'eau, fixé à 4,85 ppm à 298 K. Les expériences dite phase sensitive TOCSY [167] et NOESY [168] ont été enregistrées à 298 K en mode absorption pur [169]. La durée dite de "spin lock" au cours des expériences TOCSY était de 30 ms et les temps de "mixing" au cours des expériences NOESY se situaient dans une plage de 30 à 120 ms. Les données RMN ont été traitées sur une station de type Sun Sparc avec le logiciel Varian, en utilisant les procédures classiques de filtrage du signal résiduel de l'eau, d'amélioration de la résolution, et de correction des lignes de base [166].

### III-RÉSULTATS ET DISCUSSION

L'analyse des expériences une dimension  $^1H$  dans  $D_2O$  montre un spectre caractéristique d'une protéine paramagnétique. En effet, on peut observer, sur la Figure 4a, un certain nombre de systèmes de spin déplacés à haut et bas champ au delà de la zone diamagnétique qui se situe entre 0 et 15 ppm. Ces protons sont directement affectés par le magnétisme du fer à l'état réduit, qui est dans un état haut spin  $S=2$  [170].

L'identification de ces protons n'a pu être faite que très partiellement. Seuls les systèmes déplacés entre 40 et 60 ppm ont été identifiés comme étant les protons  $H_{N\delta}$  des résidus histidines ligands (Figure 4b). L'identification des autres systèmes déplacés n'est resté qu'à l'état d'hypothèses. Cette attribution de systèmes affectés par le paramagnétisme nécessite de connaître la position du tenseur de susceptibilité magnétique de la protéine. Or

sans structure connue, nous ne pouvions le définir que de manière arbitraire. De là, et en fonction de la localisation des protons par rapport aux axes du tenseur de susceptibilité (positionné arbitrairement), on a pu émettre des hypothèses sur le type de déplacement que l'on devrait observer pour tel ou tel proton affecté par le magnétisme du fer.

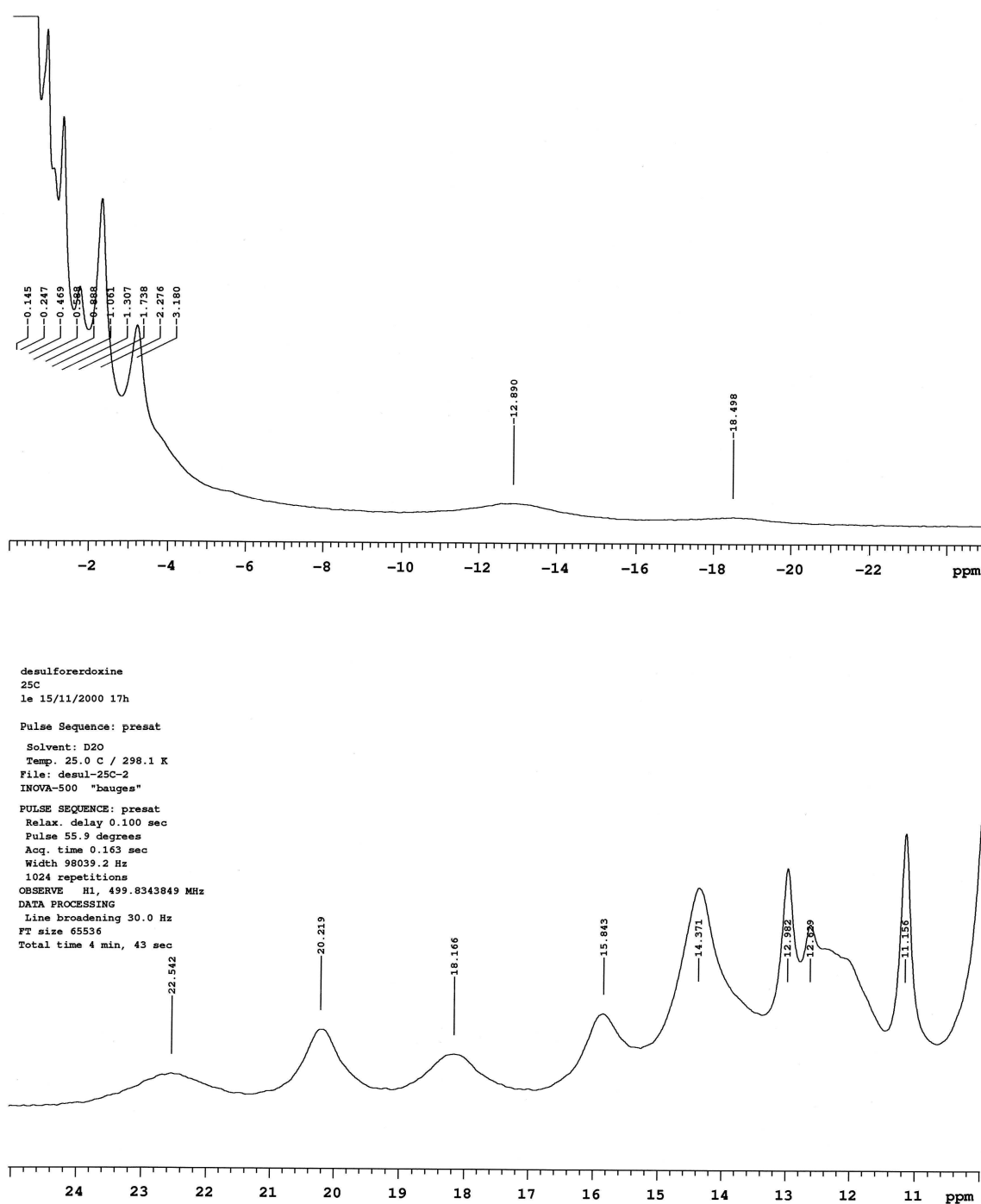
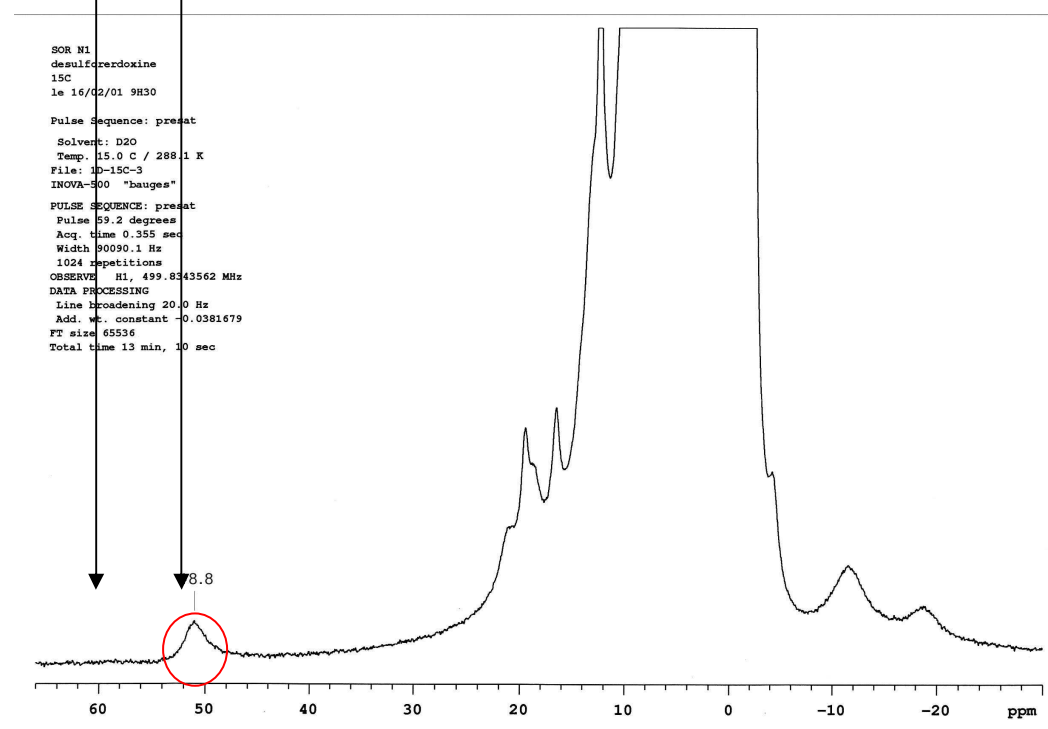
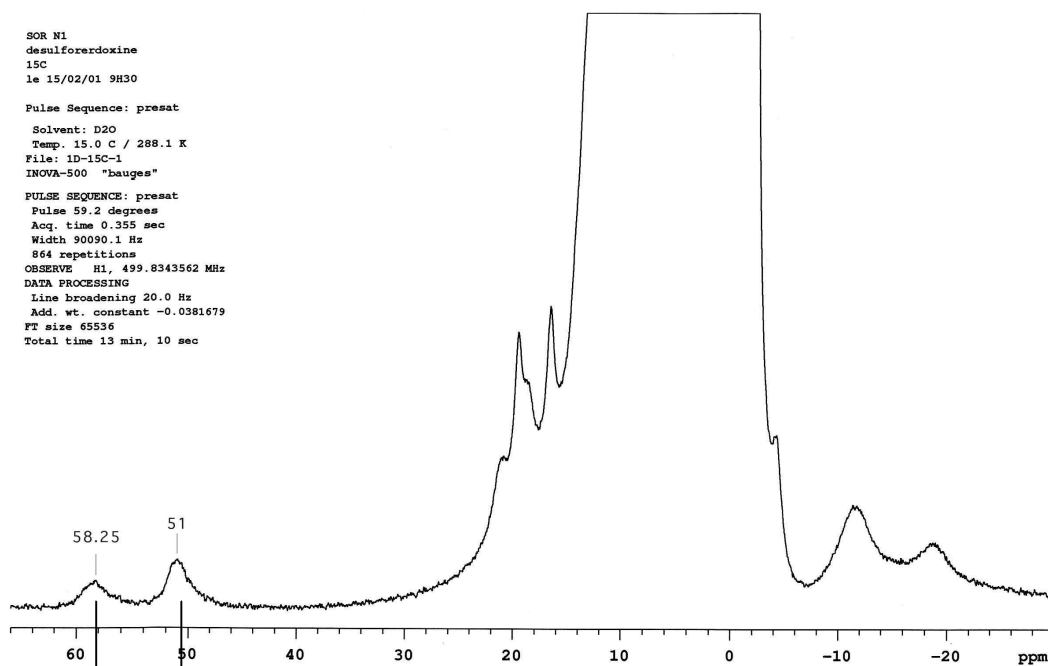


Figure 4a : Spectre 1D de la SOR de *T.p*



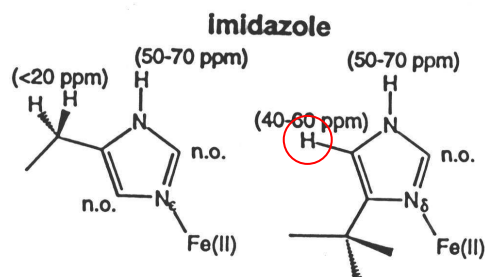


Figure 4b : Spectre 1D de la SOR de *T. pallidum*. Le signal à 51 ppm correspond à celui du H<sub>C</sub> du groupement imidazole de l'histidine 118, seule histidine ligand liée par son N<sub>δ</sub>

Parallèlement à ce travail, nous avons essayé d'analyser les expériences 2D-homo nucléaires TOCSY et NOESY afin d'effectuer une attribution des résidus de la protéine. L'idée était de repérer des systèmes de spins caractéristiques avec les différents spectres TOCSY et d'effectuer les connections séquentielles entre les résidus avec le spectre NOESY.

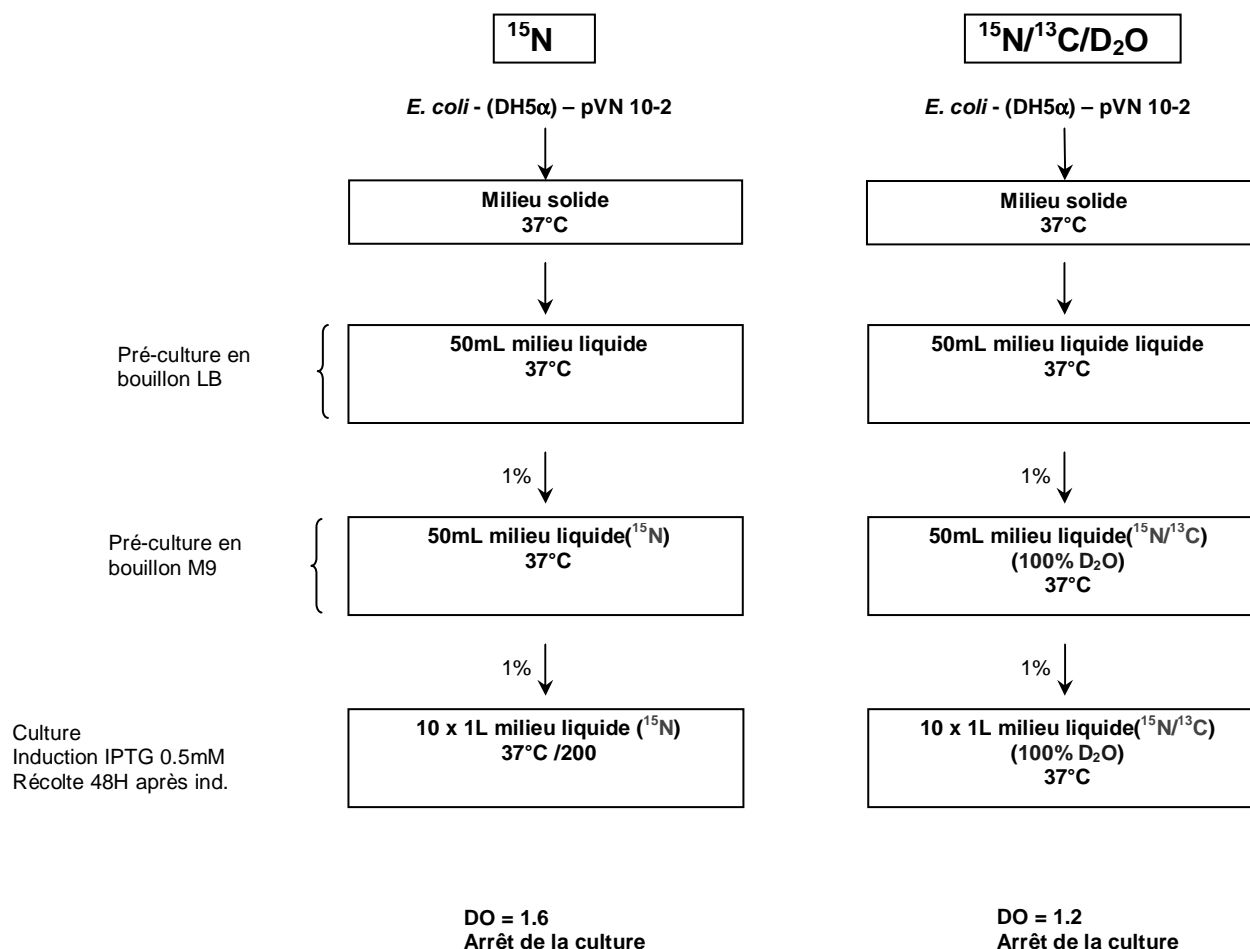
Si quelques protons ont pu être identifiés, nous n'avons jamais pu connecter sans ambiguïté les quelques résidus que nous avons identifiés. La principale difficulté venait du manque de résolution des spectres 2D, induisant trop de superposition des taches de corrélation pour attribuer les systèmes de spin. Nous avons donc décidé de modifier notre stratégie en nous adressant au laboratoire de RMN de l'Institut de Biologie Structurale (IBS) de Grenoble. À la vue des difficultés rencontrées précédemment, il a été décidé d'effectuer dans un premier temps un simple marquage isotopique  $^{15}\text{N}$ , suivi en cas de résultat positif de la première expérience, d'un triple marquage isotopique  $^{15}\text{N}/^{13}\text{C}/\text{D}$  de la protéine selon les protocoles présentés à la Figure 4.

## IV-PRODUCTION DE SOR MARQUÉE DE *T. pallidum*

### IV.1-Production de la protéine SOR marquée

Le simple marquage isotopique  $^{15}\text{N}$  et le triple marquage isotopique  $^{15}\text{N}/^{13}\text{C}/\text{D}$  de la protéine sont effectués suivant les protocoles présentés à la Figure 5.

Ces protocoles sont totalement identiques à celui présenté dans le chapitre matériels et méthodes, mis à part l'introduction des isotopes nécessaires aux différents marquages. Notons que la culture dans l'eau lourde a nécessité des temps de culture de deux à trois fois plus longs que pour la protéine non marquée.

Figure 5 : Protocole de marquage isotopique de la SOR de *T. pallidum*

La purification de la protéine est réalisée via un tamisage moléculaire sur ACA 54 suivi d'une chromatographie échangeuse d'anions sur résine UnoQ. La pureté de la protéine, analysée après chaque étape par SDS-PAGE est supérieure à 90% après la dernière étape sur UnoQ.

## V-RÉSULTATS DE PRODUCTION DE LA PROTÉINE SOR

1) L'analyse par SDS-PAGE des extraits cellulaires induits d'*E. coli* a montré une expression importante d'un polypeptide d'environ 16 kDa, absent dans les extraits cellulaires non induits. Le séquençage N-terminal de cette bande à 16 kDa a révélé qu'il s'agit d'un polypeptide (GRELSFFLQK) dont la séquence correspond au produit du gène de la SOR de *T. pallidum* sans la méthionine N-terminale.



2) La pureté de la protéine, analysée après chaque étape par SDS-PAGE (Figure 1) est supérieure à 90% après la dernière étape sur UnoQ.

3) Bilan de la purification de la SOR de *Treponema pallidum* à partir de 10 litres de culture :

	Quantité (mg) de protéines	DO <sub>280</sub> /DO <sub>644</sub>	Rendement <sup>a</sup> (%)
Extraits solubles totaux	1370	—	—
ACA 54	59	17	4.3
UnoQ	<b>21</b>	8	1.5

(a) Rendement =  $[Q^{16}(\text{mg}) \text{ protéines avant purification} / Q^{16}(\text{mg}) \text{ protéines après purification}] \times 100$

Tableau 1 : bilan de purification de la SOR de *T. pallidum*. On peut constater que les rendements sont faibles. Toutefois la quantité de protéine produite est suffisante pour produire un échantillon RMN.

## V.1-Caractérisation de la SOR de *Treponema pallidum*

### V1.1-Caractéristique biochimique

L'analyse de la SOR purifiée, simplement marquée <sup>15</sup>N, par spectrométrie de masse en mode électrospray et en conditions dénaturantes, a relevé la présence majoritaire d'une chaîne polypeptidique de 13964 Da, correspondant à la masse attendue sans la méthionine N-terminale et la présence minoritaire d'une chaîne polypeptidique de 13832 Da correspondant à la masse de la protéine avec la méthionine N-terminale (164 <sup>15</sup>N, donc +164 Da).

Par contre l'analyse de l'échantillon de SOR triplement marqué révèle la présence d'une seule entité correspondant à la masse attendue de la protéine triplement marquée avec la méthionine N-terminale.

La structure oligomérique de la protéine a été analysée par gel filtration et indique qu'il s'agit d'une protéine de 28kDa, suggérant une structure homodimérique de la SOR de *Treponema pallidum*.

La teneur en métaux (Fe, Mn, Zn) a été analysée par absorption atomique, et la concentration en protéine a été dosée par la méthode de Bradford. La protéine purifiée contient 0.7 Fe / chaîne polypeptidique et aucune trace de Mn ni de Zn n'a été détecté.

## V.1.2-Caractéristiques spectroscopiques

### V.1.2.1-Spectroscopie UV-visible

La protéine recombinante purifiée présente deux faibles bandes d'absorptions centrées à 644 et 330 nm (Figure 6), caractéristique du seul centre à Fer de type  $[\text{Fe}^{\text{II}}(\text{N}_{\text{His}})_4(\text{S}_{\text{Cys}})]$ , partiellement oxydé<sup>[171]</sup>.

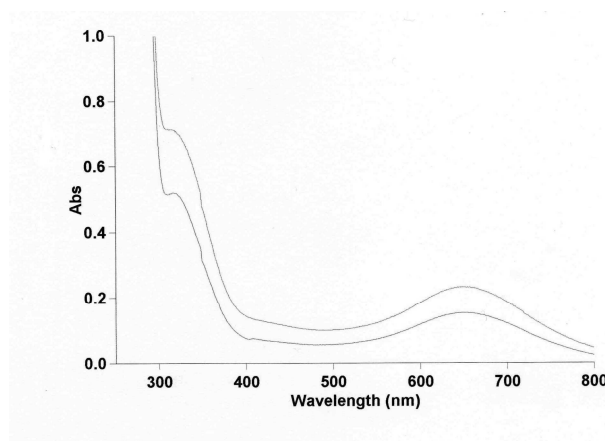


Figure 6 : Spectre UV-vis de la SOR de *T. pallidum* purifiée, oxydée.

La protéine recombinante traitée par un équivalent de ferricyanure devient bleu verte, et l'intensité des deux bandes d'adsorption à 644 nm et 330 nm augmentent. Celle à 644 nm est une bande LMCT  $\text{S} \rightarrow \text{Fe}(\text{III})$  du centre II. La valeur du coefficient d'extinction molaire déterminé à 644 nm,  $\epsilon_{644} = 2300 \text{ M}^{-1} \text{ cm}^{-1}$ , permet de déterminer la concentration du site à Fer de la protéine, sous forme complètement oxydée.

### V.1.2.2-Spectroscopie de masse

L'analyse de la SOR purifiée par spectrométrie de masse en mode électrospray et en conditions dénaturantes, a relevé la présence majoritaire d'une chaîne polypeptidique de 15146 Da, correspondant à la masse attendue avec la méthionine N-terminale (Figure 7).

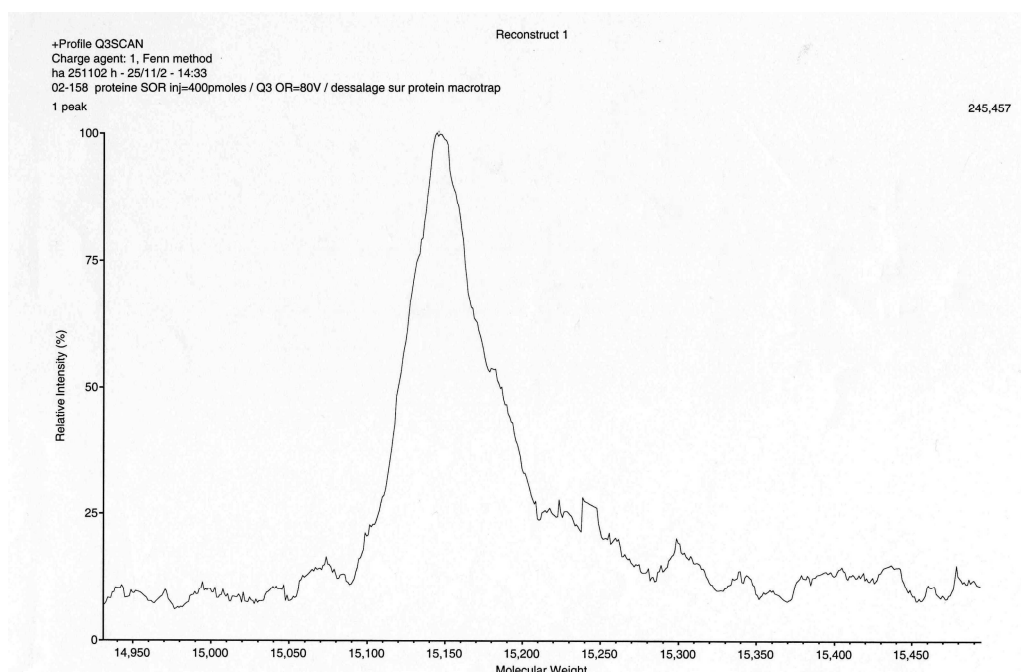


Figure 7 : Spectre de masse (électrospray reconstitué) de la SOR de *T. pallidum* purifiée, oxydée

La structure oligomérique de la protéine a été analysée par gel filtration (ACA54 de chez ) et indique qu'il s'agit d'une protéine de 2x14 kDa, suggérant une structure homodimérique de la SufA d'*E. coli*.

## VI-ANALYSE RMN

La protéine SOR est dissoute dans un tampon phosphate à 50 mM de pH égal à 6.9 auquel on ajoute 10% en volume de D<sub>2</sub>O, nécessaire au "lock" du signal RMN. On ajoute un cocktail antiprotéases de type Complete<sup>TM</sup> sans EDTA (Boehringer) afin de se prémunir de l'action d'éventuelles protéases qui pourrait toujours être présentes à l'état de trace. La protéine est ensuite réduite selon le procédé décrit précédemment.

Le spectre <sup>15</sup>N-<sup>1</sup>H HSQC (Figure 8), enregistré à partir de l'échantillon simplement marqué <sup>15</sup>N, présente les caractéristiques d'une protéine structurée, avec un spectre relativement bien étalé dans la gamme de déplacement chimique des protons amidiques. Toutefois, seules 89 corrélations, sur les 120 attendues des NH de chacun des résidus de la protéine hors ceux des prolines, sont observées. Le magnétisme du fer affectant les ligands ainsi qu'une quinzaine de résidus dans une sphère d'environ 6 Å autour du site métallique

d'une part, et un certain nombre de superpositions d'autre part peuvent expliquer cette perte de corrélations.

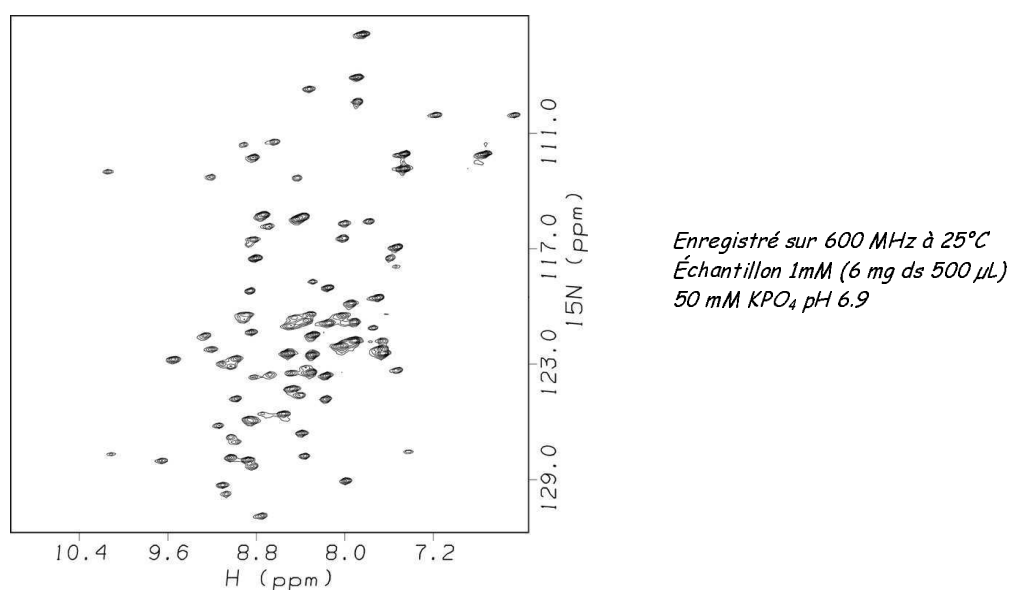


Figure 8 : Spectre  $^{15}\text{N}$ - $^1\text{H}$  HSQC de la SOR de *T. pallidum*

Suite à ce premier résultat, nous avons entrepris l'enregistrement d'expériences de triple résonance sur la protéine triplement marquée afin d'effectuer l'attribution complète de la protéine. Au préalable, pour s'assurer de la bonne intégrité de l'échantillon RMN, un spectre  $^{15}\text{N}$ - $^1\text{H}$  HSQC a été enregistré sur le spectromètre 800 MHz et est présenté sur la Figure 9.

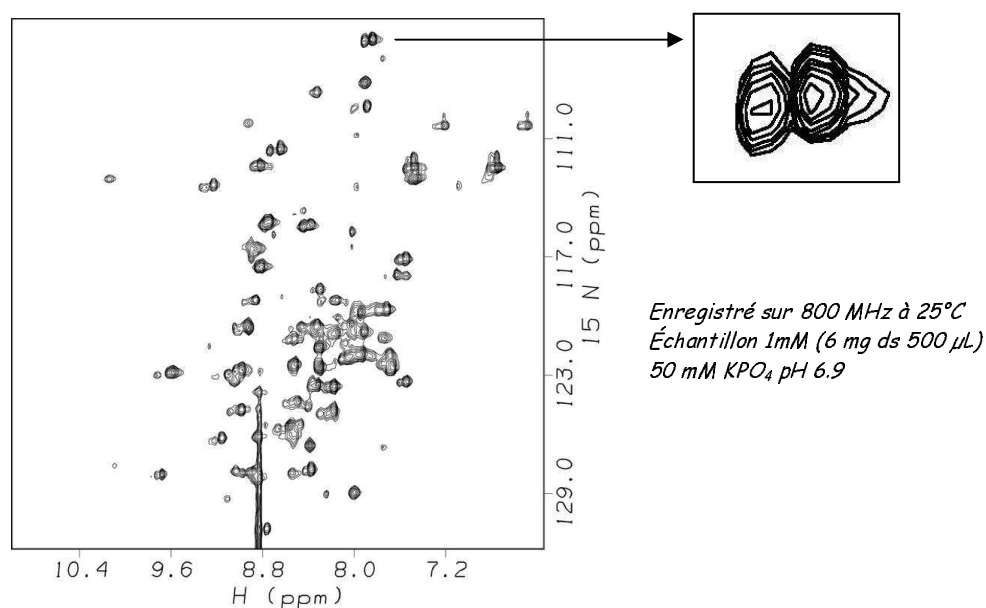


Figure 9 : Spectre  $^{15}\text{N}$ - $^1\text{H}$  HSQC de la SOR de *T. pallidum* triplement marquée, deutérée.

Si on compare ce spectre avec celui de la protéine simplement marquée  $^{15}\text{N}$ , on constate que la protéine présente globalement la même empreinte et, par conséquent, qu'elle possède la même structure. Toutefois, le décompte de corrélation est perturbé par un dédoublement des systèmes de spin. Il est par contre clair que treize corrélations ont disparues, certainement dû au fait que la culture pour cet échantillon a été effectuée dans de l'eau lourde. En effet, dans ces conditions les deutérons qui se situent dans les poches hydrophobes de la protéine ou engagés dans des liaisons hydrogènes de structures secondaires, ne s'échangent pas ou très lentement avec les protons de solvant final. Ainsi, ils ne sont pas observables avec cette expérience.

Le dédoublement de l'ensemble des taches de corrélation semble indiquer que l'on est en présence de deux conformations de la protéine avec toutefois une forme majoritaire, de l'ordre de 70% d'après le rapport d'intensité des corrélations. Nous avons éliminé un certain nombre d'hypothèses, qui auraient pu expliquer ce phénomène, parmi lesquelles :

1) aucun clivage de la méthionine terminale n'est observé sur le spectre de masse qui ne présente qu'un seul pic (Figure 7) correspondant à l'apoprotéine complète.

2) l'analyse par chromatographie sur gel filtration (Superdex75 HR 10/30 de chez Amersham pharmacia) nous montre que nous ne sommes pas en présence d'un mélange de formes oligomères mais d'un dimère (Figure 10). En effet un seul pic de protéine est élué à un poids moléculaire de  $\sim 29700\text{Da}$ , correspondant bien à 2 fois la masse du monomère.

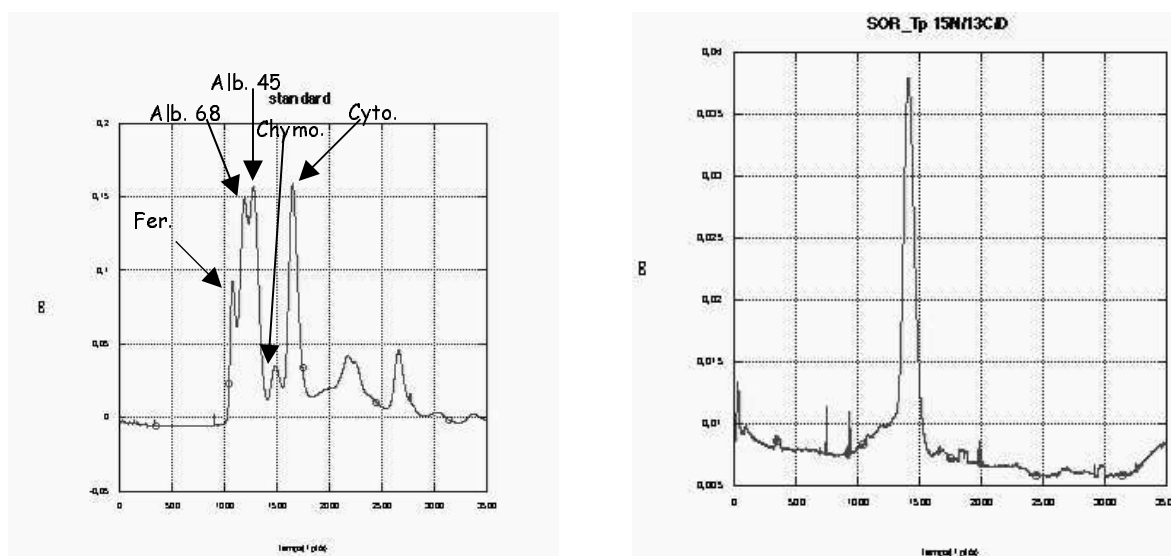


Figure 10 : Analyse de la conformation quaternaire de la SOR de *T. pallidum*. Le chromatogramme de gauche correspond aux protéines étalons (Albumine, Chymotrypsine, Albumine, Albumine et ferritine de chez Amersham). Le chromatogramme de droite à celui de la SOR.

	Vr (mL)	K <sub>AV</sub>	Ln PM
Alb. 68KDa	9.9	0.056	11.13
Alb. 45KDa	10.6	0.1	10.71
Chymo. 25 KDa	12.4	0.212	10.12
Cyto. 12.5 KDa	13.8	0.3	9.43

Tableau 2 : Calibration de la colonne Superdex 75. Vr = volume d'élution de chaque protéine ; K<sub>av</sub> =  $(V_r - V_0)/(V_c - V_0)$  avec V<sub>0</sub> égal à 9 mL (Ferritine) et V<sub>c</sub> égal à 25 mL.

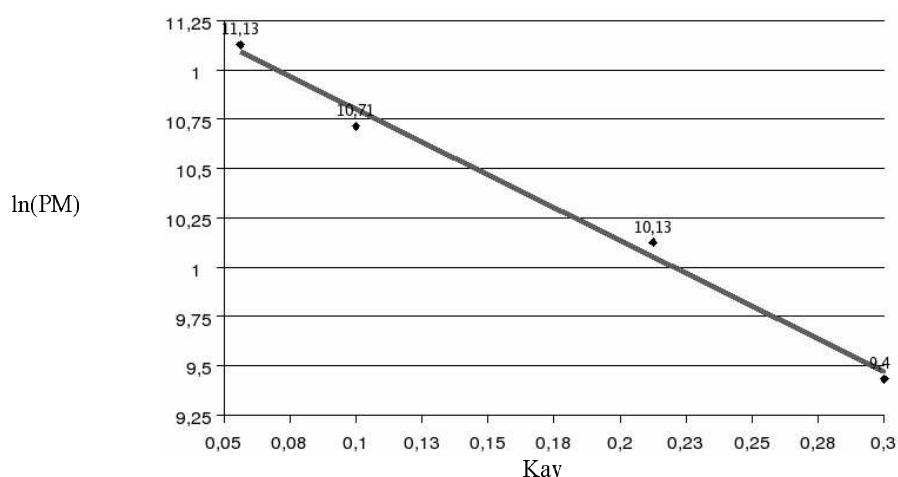


Figure 11 : Droite d'étalonnage de la colonne de gel filtration Superdex 75 H/R ; régression.

$y = -6.643x + 11.461$  avec  $r^2 = 0.99$ .

L'analyse de ces données nous permet de déterminer le poids moléculaire de l'entité injectée sur la colonne :

$$K_{AV} = (V_r - V_0)/(V_c/V_0) \quad \text{avec } V_0 \text{ égal à 9 mL (Ferritine)} \\ \text{et } V_c \text{ égal à 25 mL}$$

$$V_{r \text{ SOR}} = 11.8 \text{ mL d'où } K_{AV} = 0.175 \text{ et par conséquent } \underline{PM_{\text{SOR}} \sim 2 \times 14 \text{ kDa.}}$$

3) il est peu probable que cela soit dû à une isomérisation *cis*  $\Leftrightarrow$  *trans* de résidus proline qui ne devrait affecter que quelques résidus autour de la proline.

Malgré la présence du dédoublement des taches de corrélation, nous avons décidé d'entreprendre l'enregistrement des expériences de triples résonances pour l'attribution de la protéine. Une première série a été enregistrée avec comme expérience de base :

- $^1\text{H}$ - $^{15}\text{N}$ HSQC	$\text{H}_{\text{Ni}}$ , $\text{Ni}$
- 3D-HNCA	$\text{H}_{\text{Ni}}$ , $\text{Ni}$ , $\text{C}\alpha\text{i}-1$ , $\text{C}\alpha\text{i}$
- 3D-HNCO	$\text{H}_{\text{Ni}}$ , $\text{Ni}$ , $\text{COi}-1$
- 3D $\text{Mq}$ -HNCOCA	$\text{COi}-1$ , $\text{C}\alpha\text{i}-1$
- 3D -HN(CA)CO	$\text{COi}$
- 3D -HN(CA)CB	$\text{C}\beta\text{i}$ , $\text{C}\beta\text{i}-1$
- 3D -HN(COCA)CB	$\text{C}\beta\text{i}$

L'ensemble de ces expériences ont été décrites au chapitre II ainsi que la stratégie d'attribution qui est la même. Toutefois, on peut noter l'utilisation d'une expérience de triple résonance HNCOCA de dimensionnalité réduite, développée au laboratoire RMN de l'IBS [172], qui peut permettre de lever plus facilement des ambiguïtés d'attribution.

Le traitement de ces spectres a été effectué avec le logiciel FELIX sous station O<sub>2</sub> SGI. L'analyse des spectres s'est avérée extrêmement délicate (Figure 11).

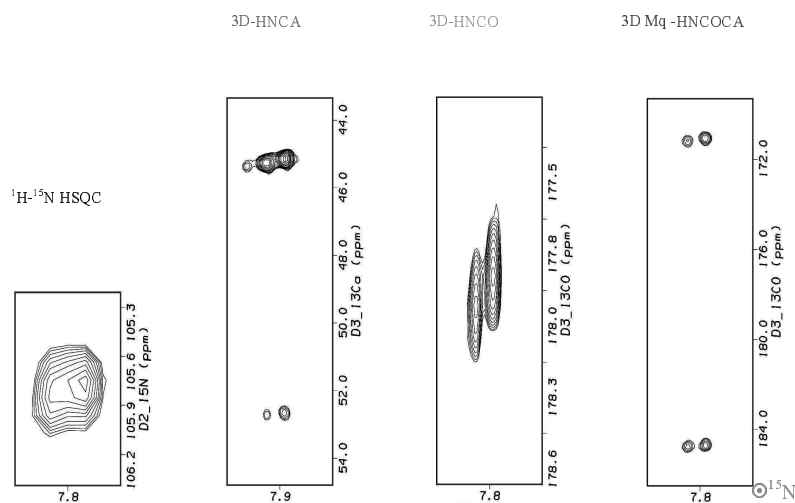


Figure 11 : Spectres de SOR Tp  $^{15}\text{N}$  /  $^{13}\text{C}$  / 75% D (1mM) enregistrés sur un spectromètre Varian 800 MHz et à 25°C (50 mM phosphate de potassium pH 6.9).

En effet nous sommes en présence d'une protéine paramagnétique dimérique qui présente un dédoublement de ses systèmes de spin. A ce stade, il a bien fallu se rendre à l'évidence que la détermination de la structure de la SOR de *T. pallidum* serait un travail de très longue haleine à cause des ambiguïtés d'attribution dues aux nombreuses superpositions dans les coupes des spectres 3D. Dans le cadre de la thèse, il a été décidé de ne pas poursuivre cette étude. Cependant il est envisagé de reprendre cette étude, après l'expérience acquise sur SufA, même si une structure cristallographique de la SOR de *T.p* a été publiée entre temps [173]. Cette structure pourrait servir de base de travail pour d'éventuelles expériences RMN (dynamique, interaction avec des partenaires, caractérisation de mutants dirigé). D'autre part il serait tout aussi intéressant de comprendre à quoi peut être dû le dédoublement observé. Cela pourrait par exemple refléter la présence de deux conformations de la protéine en échange lent, des conformations différentes du site actif, ou bien une tautométrie au niveau des  $\text{N}_\delta$  et  $\text{N}_\epsilon$  du cycle des histidines ligands. En effet au cours de la dernière étape de purification de la protéine sur colonne échangeuse d'anion UNOQ, nous avons remarqué la présence de deux entités. Celles-ci semblent être en équilibre puisqu'après avoir été isolées, si on réanalyse chacune d'entre elles sur colonne UNOQ on retrouve à nouveau deux formes. Nous n'avons par contre pas déterminé la cinétique de cet échange et sa nature.



## Production de la protéine SOR de *T. pallidum*

### I-MATÉRIEL BIOLOGIQUE

#### I.1-souches et plasmides

Le gène de structure de la SOR de *Treponema pallidum* est cloné dans le plasmide d'expression pJF119EH, entre les sites EcoR1 et BamH1, et sous le contrôle du promoteur ptac, inductible à l'IPTG. Le plasmide résultant a été nommé pVN10-2. Des souches bactériennes DH5 $\alpha$  d'*E. coli* ont été transformées par le plasmide pVN10-2. La protéine recombinante a été surproduite chez *E. coli* et purifiée selon la protocole décrit ci-après.

#### I.2-Milieu de culture

Le milieu minimum M9 supplémenté en sels minéraux et en molécules organiques est utilisé pour la pré-culture et la culture des souches DH5 $\alpha$  transformées par le plasmide pVN-10-2.

#### I.3-Culture de la souche bactérienne et surexpression de la SOR

La souche DH5 $\alpha$ /pVN10-2 est étalée sur boîte de pétri LB contenant de l'ampicilline (200 $\mu$ g/ml). A partir d'une colonie, 100 mL de milieu minimum M9 contenant le même antibiotique sontensemencés et cultivés en aérobiose à 37°C pendant 16 heures. Cette pré-

culture sert alors à ensemercer, à 1 %, 10 L de milieu minimum M9 contenant le même antibiotique et 1 mM de FeCl<sub>3</sub>. Lorsque les cellules sont en phase exponentielle ( $DO_{600} = 0.3$ ), la surexpression de la SOR est induite par addition d'IPTG (1.5 mM final). La culture se poursuit sur la nuit pour ensuite récolter les cellules par centrifugations à 4000 rpm, à 4°C, pendant 45 minutes.

## II-TECHNIQUES DE BIOCHIMIE

### II.1-Préparation d'extraits cellulaires

Le culot bactérien est repris dans 3 volumes (volume par rapport à la masse cellulaire) de tampon 100 mM Tris-HCl, pH 7.6, en présence de 0.2 mg/ml d'inhibiteurs de protéases (Complete<sup>TM</sup> de Boehringer). La suspension bactérienne est alors soumise aux ultrasons (amplitude 20%, 4°C, 5 minutes) afin de lyser les cellules. Les débris cellulaires (membranes externes et cellules non lysées) sont éliminés par ultracentrifugation à 45000 rpm (rotor Beckman TI50), à 4°C, pendant 90 minutes. On obtient ainsi les extraits solubles.

### II.1-Traitement des extraits solubles

Afin de précipiter les acides nucléiques, les extraits solubles sont traités par du sulfate de streptomycine (solubilisé dans du tampon Tris-HCl 50 mM pH 7.6) à une concentration finale de 2% (poids/volume) pendant 1 heure sous agitation à 4°C. Après centrifugation à 15000 rpm, à 4°C, pendant 20 minutes, le surnageant est traité par du sulfate d'ammonium à 70% (poids/volume), pendant 1 heure à 4°C, afin de précipiter toutes les protéines. Après centrifugation à 15000 rpm, à 4°C, pendant 20 minutes, le culot protéique peut être soit congelé dans l'azote liquide et stocké à - 80°C, soit directement utilisé.

## **II.3-Purification des protéines**

### **II.3.1-Chromatographie par tamisage moléculaire sur gel ACA 54**

Après préparation des extraits cellulaires solubles, le culot en sulfate d'ammonium est remis en suspension avec un volume de tampon A (Tris-HCl 25 mM, pH 7.6) pour atteindre une concentration finale en protéines d'environ 70mg/ml. 12 mL de cette solution protéique (soit 840 mg) sont centrifugés à 15000 rpm, à 4°C, pendant 10 min, puis déposés sur le gel ACA 54 (2.6x70 cm, 372 mL, Ultrogel, IBF) préalablement équilibré par 3 volumes de tampon A. Cette colonne ACA 54 sépare de façon optimale les protéines dont la masse moléculaire est comprise entre 5 et 70 kDa. L'élution est réalisée à un débit de 0.3 mL/min et des fractions de 4 mL sont collectées. Celles-ci sont analysées par spectroscopie électronique UV-visible : les fractions correspondant à la SOR sont rassemblées en fonction de leur absorbance à 650 nm (longueur d'onde caractéristique de centre FeN<sub>4</sub>S à l'état oxydé). Les fractions rassemblées sont alors diluées avec 5 volumes de tampon B (Tris-HCL 10 mM, pH 7.6) et concentrées par ultrafiltration sur cellule Diaflo équipée d'une membrane YM10 (Amicon co.).

### **II.3.2-Chromatographie par résine échangeuse d'anions Uno Q**

Les fractions lavées et concentrées sont déposées à un débit de 0.5 mL/min sur une colonne UnoQ-6 (Bio Rad, 6 mL) préalablement équilibrée par 5 volumes de tampon B. La colonne est alors lavée à un débit de 1.5 mL/min avec 10 volumes de tampon B. L'élution des protéines est réalisée au même débit par deux gradients linéaires, le premier allant de 0 à 150 mM NaCl en 10 volumes colonnes, le second allant de 150 à 500 mM en 2 volumes colonnes. Les fractions de 1.5 mL sont analysées par spectroscopie électronique UV-vis de la même façon que décrit précédemment.

## **III-ANALYSES BIOCHIMIQUES**

### **III.1-Electrophorèse de protéines : SDS-PAGE**

L'électrophorèse sur gel de polyacrylamide, en conditions dénaturantes, permet de séparer les protéines en fonction de leur masse moléculaires, et ainsi d'estimer la masse moléculaire de leur(s) chaîne(s) polypeptidiques ainsi que leur degré de pureté. Elle est également utilisée avec des cellules entières afin de contrôler la surexpression d'une protéine d'intérêt.

### III.1.1-Préparation des échantillons

Les échantillons protéiques sont dilués dans 2 à 6 volumes de tampon de dénaturation (60 mM Tris-HCl pH 6.8, 2% SDS, 25% glycérol, 0.2% bleu de bromophénol et 0 à 5%  $\beta$ -mercaptoéthanol). Afin de compléter la dénaturation, ils sont chauffés à 95°C pendant 5 min. Après avoir été refroidi, les échantillons dénaturés sont centrifugés à 15000 rpm durant 5 min, puis le culot protéique est remis en suspension dans quelques  $\mu$ L de tampon de dénaturation.

Concernant les cellules, 1 mL de culture est centrifugée à 15000 rpm durant 2 min, puis le culot bactérien est repris dans du tampon de dénaturation et chauffé à 95°C durant 5 min.

Après chauffage, tous les échantillons sont centrifugés à 15000 rpm, pendant 5 min puis déposés sur le gel.

### III.1.2-Conditions de migration et révélation des protéines

La composition des gels de polyacrylamide utilisés est indiquée dans le tableau 1. La migration est effectuée sous une tension de 150 V, à 30 mA pendant 60 min, dans le tampon de migration (3g/L Tris-base, 14.4 g/L glycine, 1g/L SDS).

	Gel de séparation	Gel de concentration
Polyacrylamide	15%	5%
Acrylamide 30%	3 mL	0.85 mL
Bisacrylamide 8% (w /v)		
Tris-HCl 1M pH 8.8	1.8 mL	0.78 mL
Glycérol 50%	1.2 mL	0 mL
H <sub>2</sub> O	0 mL	3.4 mL
SDS 10%	60 $\mu$ L	50 $\mu$ L

APS (100mg/mL)	30 $\mu$ L	25 $\mu$ L
TEMED	4 $\mu$ L	4 $\mu$ L

Tableau 1 : Composition des gels de polyacrylamide

Après migration, le gel est coloré pendant une heure avec une solution de bleu de Coomassie (0.25% bleu de Coomassie, 45% éthanol, 10% acide acétique, 10% glycérol). Puis la décoloration est effectuée dans une solution contenant 30% d'éthanol, 10% d'acide acétique et 5% de glycérol. Les marqueurs de masse moléculaire utilisés sont la phosphorylase (94kDa), la sérum albumine bovine (67kDa), l'ovalbumine (43kDa), l'anhydrase carbonique (30 kDa), l'inhibiteur de trypsine (20.1 kDa) et l' $\alpha$ -lactalbumine (14.4 kDa). (kit Low Molecular Weights de Pharmacia).

### III.2-Dosage standard

Les protéines solubles sont dosées selon le protocole décrit par Bradford [<sup>174</sup>]. Cette méthode est basée sur le changement de coloration de bleu de Coomassie qui passe du rouge ( $\lambda_{\max}$ =465 nm) au bleu ( $\lambda_{\max}$ =595 nm). Le sérum albumine de bœuf sert à la réalisation de la gamme étalon.

## Surexpression et purification de la protéine SufA

### I-MATÉRIEL BIOLOGIQUE

#### I.1-Clonage du gène SufA et construction du plasmide d'expression

Le plasmide pET/SufA, utilisé pour l'expression de la protéine SufA<sub>his</sub>, est dérivé du vecteur pET22b(+) (Novagen) (Figure 1). Ce dernier permet d'exprimer une étiquette histidine en C-terminal de la protéine afin de réaliser une purification sur colonne d'affinité. Le gène codant pour la protéine SufA est amplifié par PCR à partir de l'ADN génomique d'*E. coli* selon le protocole décrit par Ollagnier de Choudens *et al.*, 2004. Les amorces utilisées pour l'amplification de *sufA* sont les suivantes : 5'-CCGCATATGGACATGCATTCAGGAACC-3' (amorce « forward », les bases soulignées indiquent site de restriction *NdeI*) et 5'-GTTCTCGAGTACCCCAAAGCTTTCGCCACAGCC-3' (amorce « reverse », les bases soulignées indiquent le site *XhoI*). L'amplification PCR est réalisée comme suit : dénaturation de ADN génomique 10 minutes à 94°C ; addition de l'ADN polymérase *Pwo* (0,5 unit), du mélange de désoxynucléotides (0,2 mM de chaque) et des amorces (0,5 mM). Le programme d'amplification inclut trente cycles (30s à 94°C, 30s à 50°C et 30s à 72°C) et une phase d'élongation finale de 10 minutes à 72°C. Le produit PCR est ensuite digéré par les enzymes *NdeI* et *XhoI* puis cloné dans le vecteur pET22b(+) préalablement digéré avec les mêmes enzymes de restriction. Le gène cloné fût séquencé afin de contrôler qu'aucune erreur n'ait été introduite pendant la réaction d'amplification. Le plasmide ainsi obtenu est nommé pET/SufA.

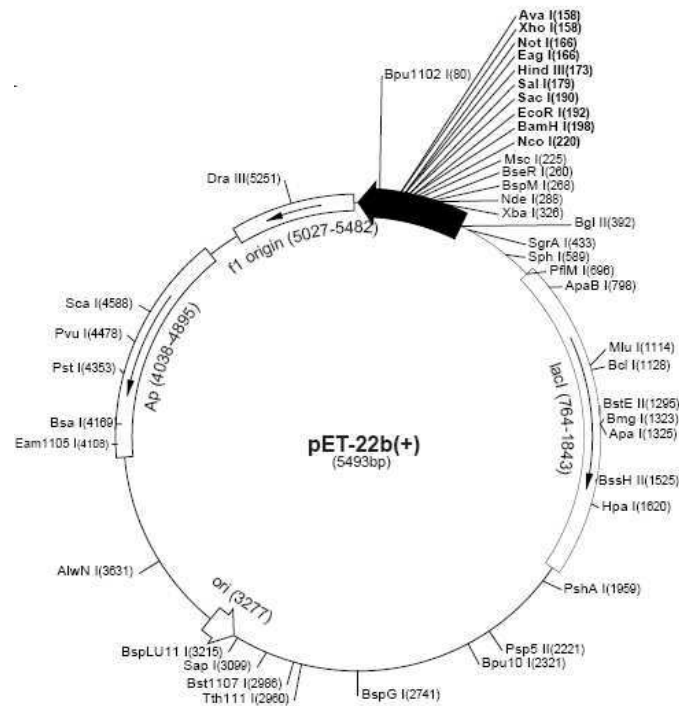


Figure 1 : vecteur pET-22b(+)

## II-PRODUCTION ET PURIFICATION DE LA PROTÉINE SufA SIMPLEMENT ET DOUBLEMENT MARQUÉE PAR LES ISOTOPES $^{15}\text{N}/^{13}\text{C}$

### II.1-Production

Les cellules BL21(DE3) compétentes *E. coli* sont transformées avec le vecteur pET/SufA. Une première pré-culture en milieu riche LB est réalisée suivi d'une pré-culture en milieu M9 (2g/L de  $^{13}\text{C}_6$ -Glucose, 1 g/L de  $^{15}\text{NH}_4\text{Cl}$ , 1 mM  $\text{MgSO}_4$ , 0,1 mM  $\text{CaCl}_2$ , 2 mg/L thiamine) est réalisée. Le milieu est complété avec de l'ampicilline (100  $\mu\text{g/mL}$ ) de manière à maintenir le plasmide pET/SufA dans la bactérie. Une culture de 3 L en milieu M9 est lancée à partir de la pré-culture en milieu M9 marqué. Lorsque la culture atteint une  $\text{DO}_{600\text{nm}}$  de 0,5, la production de la protéine est induite pendant 3 à 4 heures par addition d'une solution d'Isopropyl- $\beta$ -D-thiogalactopyranoside (Eurogentec) à une concentration finale de 0,5 mM. Les milieux M9 utilisés pour la pré-culture et pour la culture sont soit simplement marqués par  $^{15}\text{N}$ , soit doublement marqué par  $^{15}\text{N}/^{13}\text{C}$  (Figure 2). Le  $^{13}\text{C}_6$ -Glucose et le  $^{15}\text{NH}_4\text{Cl}$  sont respectivement les seules sources de carbone et d'azote.

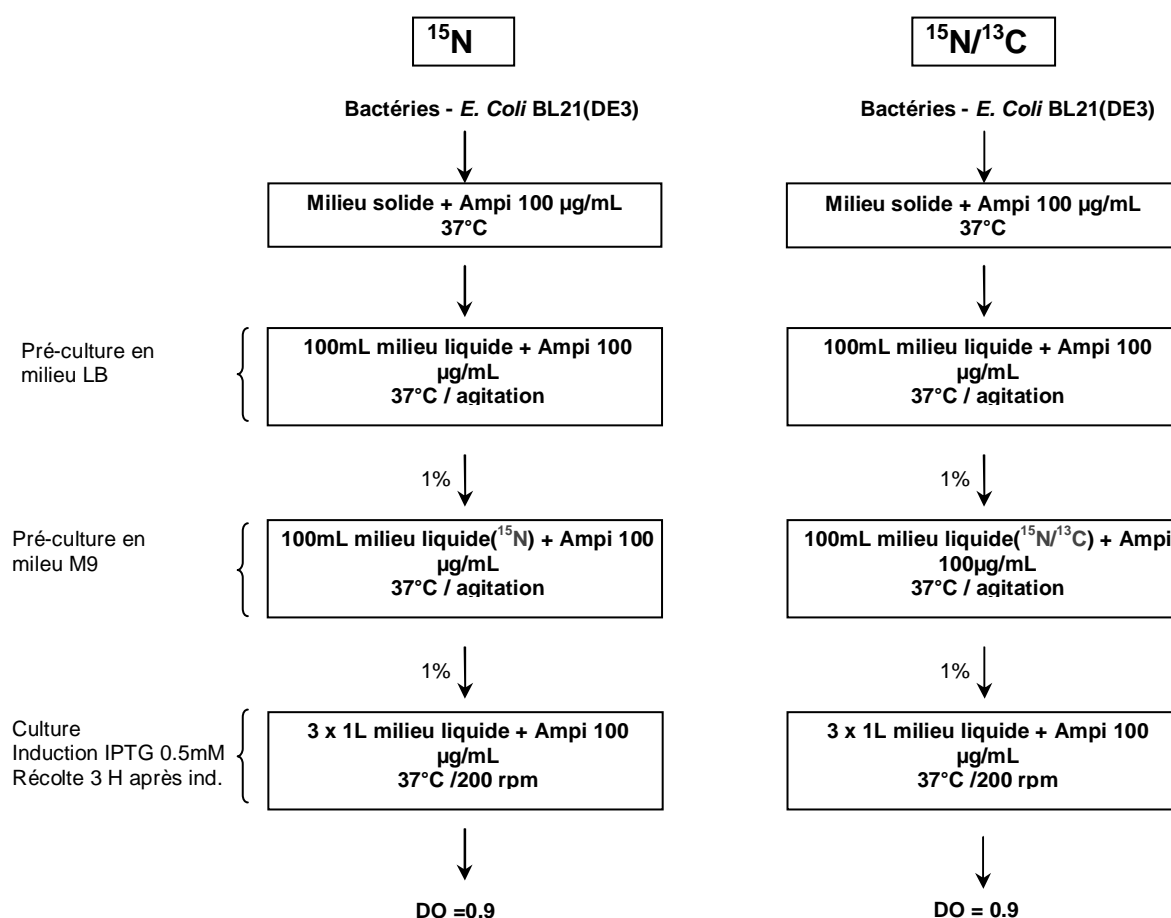


Figure 2 : production de la protéine SufA marquée

## II.2-Purification

### II.2.1-Extraction

Les culots cellulaires sont repris dans 50 mL de tampon A (100 mM Tris-HCl, pH 7.5, 50 mM NaCl) et traités par sonication afin de casser la paroi des bactéries. Le lysat cellulaire est centrifugé à 45000 rpm durant 90 min, à 4°C. Afin de précipiter les acides nucléiques, les extraits solubles sont traités par du sulfate de streptomycine (solubilisé dans du tampon Tris-HCl 50 mM pH 7.6) à une concentration finale de 2% (p/v) pendant 1 heure sous vive agitation à 4°C. Après centrifugation à 15000 rpm, à 4°C, pendant 20 minutes, le culot protéique peut être soit directement utilisé, soit congelé dans l'azote liquide et stocké à – 80°C.



### II.2.2-Purification

Les protéines solubles sont chargées sur une colonne d'affinité NiNTA (Superflow<sup>TM</sup> de chez Quiagen) préalablement chargé avec du nickel et équilibré avec le tampon A (10 volumes de colonne). La colonne est ensuite lavée avec le tampon A (20 volumes de colonne). L'élution est effectuée avec un gradient d'imidazole (tampon A + 100 mM d'imidazole). Les fractions, analysées sur gel SDS page, contenant la protéine sont rassemblées et concentrées sur cellule Diaflo équipée d'une membrane YM10 (Amicon co.). L'échantillon ainsi obtenu (volume < 10 mL pour une concentration maximale de protéine égal à 100 mg) est ensuite déposé sur colonne de tamisage moléculaire Superdex 75 (volume 120 mL) qui a été préalablement équilibrée avec le tampon A auquel on ajoute 5 mM de DTT (3 à 4 volumes de colonne). L'élution est effectuée avec le même tampon à un débit de 0.8 mL/min. La sortie de protéine est suivie par absorption UV-vis à 260 nm. Les fractions analysées sur gel SDS et contenant la protéine SufA sont rassemblées, concentrées sur cellule Diaflo et stockées à -80°C.

### II.3-Reconstitution

SufA sous la forme apo (10 mg) est traitée en anaérobiose avec 5 mM de DTT pendant 15 min. Ensuite, ajout de 4 équivalents molaire de fer ferreux (sel de Mohr) et de soufre (Na<sub>2</sub>S) et incubation durant 3 heures à 18°C. La protéine est ensuite traitée avec 2 mM d'EDTA afin d'éliminer l'excès de fer et la protéine est chargée sur une colonne G25 qui a été préalablement équilibrée avec du tampon A. Le tampon de l'échantillon est ensuite échangé contre du tampon B (50 mM de tampon phosphate pH 7). La protéine est enfin concentrée jusqu'à 1 mM.

## Spectroscopie RMN et calcul de structure

L'ensemble des expériences RMN a été enregistré sur les spectromètres Varian de 600 et 800 MHz. L'échantillon de protéine a une concentration d'environ 1mM pour l'ensemble des expériences et est dans un tampon phosphate 50 mM de pH 6.9 en présence de DTT et de Complete<sup>TM</sup>. Les tubes sont scellés sous argon pour protéger la protéine d'éventuelles oxydations et il est conservé à 4°C entre les différentes expérience. La protéine est stable plusieurs mois dans ces conditions.

Le traitement de l'ensemble des FID a été réalisé avec l'aide du logiciel FELIX version 2001.1. Toutefois, certaines expériences ont été retraitées avec NMRPipe [138] et les spectres visualisés sous le logiciel NMRview [99]. Tous les spectres NOESY-HSQC ont été analysés avec ce logiciel.

Le calcul de structure a été effectué avec le logiciel ARIA version1.2 [128] et les structures ont été visualisées avec le logiciel MOLMOL [<sup>175</sup>]

## I-SPECTROSCOPIE RMN

*Expériences d'attribution des résonances de la chaîne principale :*

Tableau1 : Paramètres des expériences 3D d'attribution de la chaîne principale pour l'enregistrement et le traitement des données.

Expériences	<sup>15</sup> N- <sup>1</sup> H HSQC	HNCA	CBCANH	CBCA(CO)NH	HNCO	HN(CA)CO
Acquisition						
MHz	600	600	600	600	600	600
Nombre de points dans les dimensions proton, carbone et azote respectivement						
np H1	1024	512	1024	1024	1024	1024
ni C13	128	128	59	59	128	80
ni2 N15		56	47	55	55	49
Largeurs spectrales (Hz)						
sw H1	13008.13	7499.7656	7499.7656	7499.7656	7499.7656	7499.7656
sw C13		4524.57	10000	10000	2200	2200
sw N15	2000	2000	2000	2000	2000	2000
f1180	NON	(DW D2)				
f2180	NON	(DW D3)				
Traitements des données						
	Tailles des matrices en points					
D1(H)*D2(N)* D3(C)		512x128x 256	512x128x 128	512x128x128	512x128x 256	512x128x 256
cnv	Traitement de suppression de l'eau					
fenêtre d'apodisation	Fonction cosinus carré 90° (paramètre FELIX :qss 90 (sensibilité))					
	Correction des lignes de base					
	Porteuse en ppm					
H1		4.74	4.74	4.74	4.74	4.74
C13		54.87	44.87	44.87	172.89	54.87

*Expériences d'attribution des résonances des chaînes latérales :*

Tableau2 : Paramètres des expériences 3D d'attribution des chaînes latérales pour l'enregistrement et le traitement des données.

Expériences	HC(C)H- TOCSY	(H)CCH- TOCSY	H(CC)CONH- TOCSY	(H)CC(CO)NH- TOCSY	HAHB(CO)NH	
Acquisition						
MHz	600	600	600	600	600	
Nombre de points dans les dimensions proton, carbone et azote respectivement						
np H1	1024	1024	1024	1024	768	
ni C13	90	112	180	80	128	
ni2 N15	126	128	55	55	55	
Largeurs spectrales (Hz)						
sw H1	7499.76	7509.6	7509.6	7499.76	7509.6	
sw 1	7499.76	12001.2	4000	12000	3000	
sw 2	12001.2	12001.2	2000	2000	2000	
f2180	Demi well-time en dimension D2					
f1180	Demi well-time en dimension D3					
Traitements des données						
	Tailles des matrices en points					
D1(H)*D2(N)*D3(C)	512x256 x256	512x256 x256	512x128x512	512x128x256	512x128x256	
	Correction des lignes de base					
D1 (H)	OUI					
D2 (N) et D3 (C)	OUI					
	Porteuse en ppm					
H1	4.74	4.74	4.74		4.74	
C13		34.75		35	122.94	
15N			120.8	120.8		

*Expériences NOESY-HSQC :*

Tableau3 : Paramètres des expériences 3D NOESY-HSQC pour l'enregistrement et le traitement des données.

Expériences	<sup>15</sup> N NOESY-HSQC	<sup>13</sup> C NOESY-HSQC
acquisition		
MHz	800	800
Nombre de points dans les dimensions proton, carbone et azote respectivement		
np H1	1024	1280
ni C13	180	128
ni2 N15	75	60
Largeurs spectrales		
sw H1	12001.2	10000
sw 1	9000	7000
sw 2	2399.9	15000.9
f2180	Demi well-time en dimension D2	
f1180	Demi well-time en dimension D3	
Traitement des données		
D1(H)*D2(N)*D3(C)	512x128x256	1024x128x256
	Correction ligne de base	
D1 (H)	Oui	Oui
D2 (N) et D3 (C)	Oui	Oui
	Porteuse en ppm	
H1	4.74	4.74
C13		
15N	120.8	

*Expériences NOESY-HSQC aromatiques*

Tableau4 : Paramètres de l'expérience 3D NOESY-HSQC pour l'enregistrement et le traitement des données.

Expériences	<sup>13</sup> C NOESY-HSQC
acquisition	
MHz	800
Nombre de points dans les dimensions proton, carbone et azote respectivement	
np H1	
ni C13	
ni2 N15	
Largeurs spectrales	
sw H1	4003.9062
sw 1	8000
sw 2	5587.0928
f2180	Demi well-time en dimension D2
f1180	Demi well-time en dimension D3
Traitement des données	
D1(H)*D2(N)*D3(C)	
	Correction ligne de base
D1 (H)	oui
D2 (N) et D3 (C)	oui
	Porteuse en ppm
H1	4.739
C13	128.500

*Mesure des couplages dipolaires résiduels*

Tableau4 : Paramètres des expériences 3D HNCO pour l'enregistrement et le traitement des données.

Expériences	HNCO <sup>1</sup> D <sub>NH</sub> Accordéon 1.6	HNCO <sup>1</sup> D <sub>COCA</sub> E. COSY 1.0	HNCO <sup>1</sup> D <sub>CON</sub> DIPSAP 1.0
acquisition			
MHz	600	600	600
Nombre de points dans les dimensions proton, carbone et azote respectivement			
np H1	1024	1024	1024
ni C13	90	100	100
ni2 N15	49	50	50
Largeurs spectrales			
sw	9000.9	9000.9	9000.9
sw 1	2200	2200	2200
sw 2	1849.9	1849.9	1849.9
f2180	Demi well-time en dimension D2		
f1180	Demi well-time en dimension D3		
Traitement des données			
D1(H)*D2(N)*D3(C)	512x128x512	512x128x512	512x128x1024
	Correction ligne de base		
D1 (H)	Oui	Oui	Oui
D2 (N) et D3 (C)	Oui	Oui	Oui
	Porteuse en ppm		
H1	4.74	4.74	4.74
C13	174.96	174.992	174.96
15N	119.2	119.2	119.2

Ce tableau présente les paramètres des expériences 3D utilisées pour l'enregistrement des expériences HNCO en milieu isotrope.

Les mêmes expériences ont été enregistrées dans le milieu anisotrope (phages *Pfl* 20 mg/mL). Nous indiquons ci-dessous seulement les différences entre les expériences.

Expériences	HNCO $^1D_{NH}$	HNCO $^1D_{COCA}$	HNCO $^1D_{CON}$
np H1			
ni C13		104	96
ni2 N15	48		52

Remarque : la mesure des RDC nécessite d'être très précis quant à la position des pics. Pour éviter d'introduire des erreurs, il est nécessaire d'effectuer une optimisation des volumes de corrélation. Felix propose une macro assez solide qui permet d'effectuer ce travail qui se trouve dans le menu standard du logiciel. [ouvrir le menu « *optimize* » et suivre la procédure]. On peut tout aussi bien effectuer cette optimisation manuellement de manière interactive avec FELIX [ouvrir le menu « *optimize* » après avoir fait une coupe dans la troisième dimension du spectre puis suivre la procédure].

Effectuer ensuite les mesures de RDC et appliquer :

$$\begin{aligned} \text{Pour } {}^1D_{\text{NH}} \text{ et } {}^1D_{\text{COCA}} \quad & J = \text{iso\_a} - \text{iso\_b} > 0 \\ & JD = \text{aniso\_a} - \text{aniso\_b} > 0 \\ \text{Et } \mathbf{D} &= \mathbf{JD} - \mathbf{J} \end{aligned}$$

$$\begin{aligned} \text{Pour } {}^1D_{\text{CON}} \quad & J = \text{iso\_a} - \text{iso\_b} > 0 \\ & JD = \text{aniso\_a} - \text{aniso\_b} > 0 \\ \text{Et } \mathbf{D} &= \mathbf{JD} - \mathbf{J} \end{aligned}$$

ce couplage est déterminé à partir d'une séquence DIPSAP a nécessité de créer quatre matrices. Il faut donc être particulièrement attentif aux signes des couplages car on effectue de des soustractions d'un spectre par rapport à un autre, ce qui peut induire des erreurs. Toutefois ce type d'erreur est relativement vite identifié avec le logiciel MODULE qui permet d'effectuer des corrélations entre les valeurs mesurées et des valeurs calculées à partir de la structure.



*Expériences d'échange des protons amides*

Pour cette expérience, SufA a été lyophilisée puis introduite dans un tube RMN avec 450 $\mu$ L de D<sub>2</sub>O. Ensuite des expériences <sup>1</sup>H-<sup>15</sup>N HSQC ont été enregistrées sur une durée de 1023 min (17H03). 24 spectres <sup>1</sup>H-<sup>15</sup>N HSQC ont été enregistrés suivant une périodicité croissante au cours de l'expérience.

```
sufa_echange_001.ft 4 360+284 = 644s
sufa_echange_002.ft 4 644+568 =1212s
sufa_echange_003.ft 4 1212+568 =1780s
sufa_echange_004.ft 4 1780+568 =2348s
sufa_echange_005.ft 8 2348+852 =3200s
sufa_echange_006.ft 8 3200+1137=4337s
sufa_echange_007.ft 8 4337+1137=5474s
sufa_echange_008.ft 8 5474+1137=6611s
sufa_echange_009.ft 16 6611+1705=8316s
sufa_echange_010.ft 16 7916+2273=10589s
sufa_echange_011.ft 16 10189+2273=12862s
sufa_echange_012.ft 16 12462+2273=15135s
sufa_echange_013.ft 32 14735+3410=18845s
sufa_echange_014.ft 32 18045+4546=23091s
sufa_echange_015.ft 32 22591+4546=27637s
sufa_echange_016.ft 32 27137+4546=32183s
sufa_echange_017.ft 32 31683+4546=36729s
sufa_echange_018.ft 32 36229+4546=41275s
sufa_echange_019.ft 32 40775+4546=45821s
sufa_echange_020.ft 32 45321+4546=50367s
sufa_echange_021.ft 32 49867+4546=54913s
sufa_echange_022.ft 32 54413+4546=59459s
```

```
sufa_echange_023.ft: 21hrs 43m 42s -> 1303min + 38 = 1341min
```

```
sufa_echange_024.ft: 40hrs 6m -> 2406 + 38 = 2444min
```

Les intensités de pics de corrélation de chaque spectre ont été ensuite mesurées sous nmrDraw [138]. Puis on crée un fichier dans lequel on liste les intensités de tous les spectres par résidus. On effectue ensuite une interpolation des courbes d'intensité en fonction du temps pour chacun des résidus (Figure 1).

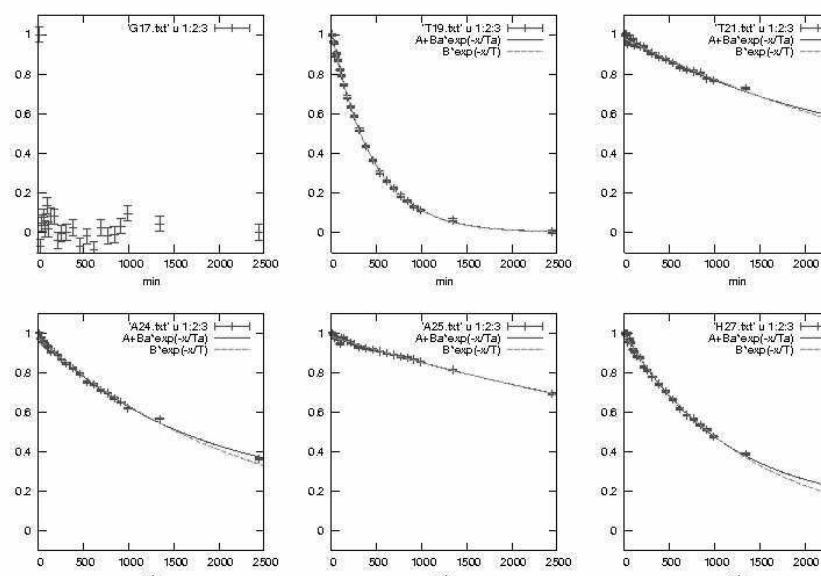


Figure 1 : Courbes d'évolution de l'intensité des pics de corrélations en fonction du temps et « fit » pour quelques résidus. On peut remarquer la différence d'évolution entre la THR 19 exposée au solvant et ALA 25 engagée dans une structure secondaire.

Des courbes d'interpolation on détermine la constante d'échange ( $k_{ex}$ ) du proton concerné. La constante d'échange extraite n'est pas utilisée directement mais est convertie en facteur de protection par rapport à un facteur de protection intrinsèque dépendant du pH. C'est ce facteur de protection qui est représenté sur la figure du paragraphe III.5.3.

## II-Calcul de structure

L'attribution automatique et le calcul des structures de la protéine SufA a été effectué avec ARIA [109, 128, 129] utilisant CNS [130]. La topologie topallhdg5.3.pro [176] et le jeu de paramètres ont été utilisés sur la base des paramètres PROSQL [177].

Les premiers calculs étaient composés de neuf itérations ARIA, les pics nOe attribués manuellement, ainsi que les contraintes TALOS, ont été utilisés pour calculer un ensemble initial de structures. Après chaque itération, les nOe ambiguës ont réattribués et recalibrés à partir des vingt structures de plus basses énergies et rejetés si ils sont violés dans plus de 50% des structures. Le paramètre « ambiguity cut-off » est réduit de 1.01 pour la première itération à la valeur finale de 0.9 à la dernière itération. La tolérance des violations est fixée à 1000 Å (pour les calculs avec la structure étendue) ou 1 Å (pour les calculs partant d'un modèle structural) pour les deux premières itérations et progressivement réduite à 0.1 Å dans la dernière itération. Pour le calcul de structure, un recuit simulé (simulated annealing SA) composé de quatre phases a été utilisé en partant d'une conformation étendue de SufA ou de la structure de la protéine homologue IscA, et en utilisant la dynamique d'angle de torsion (TAD) : i) La phase de haute température TAD est constituée de 10000 étapes à 10000 K. Celle-ci est suivie par ii) une première phase de refroidissement TAD de 8000 étapes avec une température finale de 1000 K. iii) une seconde phase de refroidissement TAD de 13000 étapes avec une température finale de 50 K. Durant le protocole de recuit simulé, les constantes de force des étapes TAD sont :

*-pour les contraintes nOE*

pour la phase haute température égales à  $10 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ , pour la première phase de refroidissement TAD égales à  $10 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$  (init) et  $50 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$  (fin), et pour la seconde phase de refroidissement TAD égales  $50 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ .

*-pour les contraintes d'angles dièdres*

pour la phase haute température égales à  $5 \text{ kcal.mol}^{-1}.\text{rad}^{-1}$ , pour la première phase de refroidissement TAD égales à  $25 \text{ kcal.mol}^{-1}.\text{rad}^{-1}$  et pour la seconde phase de refroidissement TAD égales  $200 \text{ kcal.mol}^{-1}.\text{rad}^{-1}$ .

Le nombre de structures calculées était de 100 (10) pour les itérations 0 à 7 et 200(20) pour la dernière itération, à partir des structures issues de l'itération précédente dont le nombre est indiqué entre parenthèses. Les cinquante structures de plus basse énergie ont été utilisées pour l'affinement dans l'eau.

A l'issue de ces premiers calculs les contraintes non ambiguës communes aux deux types de calcul, un avec la structure étendue et l'autre avec la structure de la protéine homologue, ont été rassemblées en un seul fichier qui a été utilisé pour les calculs suivants. Ces calculs ont été ensuite effectués uniquement à partir de la structure étendue de SufA.

La deuxième phase de calcul a donc consisté au même protocole que précédemment mais en introduisant les contraintes de liaison hydrogène, dès la première itération, ainsi que les contraintes orientationnelles RDC introduites au cours de la sixième itération.

*-pour les contraintes H-bond*

identiques aux contraintes nOe

*-pour les contraintes RDC*

*NH* : pour la phase haute température égales à 0 kcal.mol<sup>-1</sup>, pour la première phase de refroidissement TAD égales à 0.1 kcal.mol<sup>-1</sup>, et pour la seconde phase de refroidissement TAD égales 3.0 kcal.mol<sup>-1</sup>.

*COCA* : pour la phase haute température égales à 0 kcal.mol<sup>-1</sup>, pour la première phase de refroidissement TAD égales à 0.1 kcal.mol<sup>-1</sup>, et pour la seconde phase de refroidissement TAD égales 6.0 kcal.mol<sup>-1</sup>.

*CON* : pour la phase haute température égales à 0 kcal.mol<sup>-1</sup>, pour la première phase de refroidissement TAD égales à 0.1 kcal.mol<sup>-1</sup>, et pour la seconde phase de refroidissement TAD égales 6.0 kcal.mol<sup>-1</sup>.

Le nombre de structures calculées était de 100 (10) pour les itérations 0 à 7 et 100 (20) pour la dernière itération, à partir des structures issues de l'itération précédente dont le nombre est indiqué entre parenthèses. Les cinquante structures de plus basse énergie ont été utilisées pour l'affinement dans l'eau.

- <sup>1</sup> Crichton, R. R. (1990) Proteins of iron storage and transport, *Adv Protein Chem* 40, 281-363.
- <sup>2</sup> Huber, C., and Wachtershauser, G. (1997) Activated acetic acid by carbon fixation on (Fe,Ni)S under primordial conditions, *Science* 276, 245-247.
- <sup>3</sup> Beinert, H., Holm, R. H., and Munck, E. (1997) Iron-sulfur clusters: nature's modular, multipurpose structures, *Science* 277, 653-659.
- <sup>4</sup> Mortenson, L. E., Carnahan, J. E., and Valentine, R. C. (1962) An Electron Transport Factor from Clostridium Pasteurianum, *Biochemical and Biophysical Research Communications* 7, 448-&.
- <sup>5</sup> Tagawa, K., and Arnon, D. I. (1962) Ferredoxins as Electron Carriers in Photosynthesis and in Biological Production and Consumption of Hydrogen Gas, *Nature* 195, 537-&.
- <sup>6</sup> Mortenson, L. E. (1964) Purification + Analysis of Ferredoxin from Clostridium Pasteurianum, *Biochimica Et Biophysica Acta* 81, 71-&.
- <sup>7</sup> Johnson, D. C., Dean, D. R., Smith, A. D., and Johnson, M. K. (2005) Structure, function, and formation of biological iron-sulfur clusters, *Annu Rev Biochem* 74, 247-281.
- <sup>8</sup> Adman, E. T., Sieker, L. C., and Jensen, L. H. (1991) Structure of Rubredoxin from Desulfovibrio-Vulgaris at 1.5-Å Resolution, *Journal of Molecular Biology* 217, 337-352.
- <sup>9</sup> Jacobson, B. L., Chae, Y. K., Markley, J. L., Rayment, I., and Holden, H. M. (1993) Molecular-Structure of the Oxidized, Recombinant, Heterocyst 2Fe-2S Ferredoxin from Anabaena-7120 Determined to 1.7-Ångstrom Resolution, *Biochemistry* 32, 6788-6793.
- <sup>10</sup> Uhlmann, H., Beckert, V., Schwarz, D., and Bernhardt, R. (1992) Expression of Bovine Adrenodoxin in Escherichia-Coli and Site-Directed Mutagenesis of (2-Fe-2S) Cluster Ligands, *Biochemical and Biophysical Research Communications* 188, 1131-1138.
- <sup>11</sup> Golinelli, M. P., Akin, L. A., Crouse, B. R., Johnson, M. K., and Meyer, J. (1996) Cysteine ligand swapping on a deletable loop of the 2Fe-2S ferredoxin from Clostridium pasteurianum, *Biochemistry* 35, 8995-9002.
- <sup>12</sup> Holden, H. M., Jacobson, B. L., Hurley, J. K., Tollin, G., Oh, B. H., Skjeldal, L., Chae, Y. K., Cheng, H., Xia, B., and Markley, J. L. (1994) Structure-Function Studies of 2Fe-2S Ferredoxins, *Journal of Bioenergetics and Biomembranes* 26, 67-88.
- <sup>13</sup> Fontecave, M. (2006) Iron-sulfur clusters: ever-expanding roles, *Nature Chemical Biology* 2, 171-174.
- <sup>14</sup> Kent, T. A., Munck, E., Dunham, W. R., Filter, W. F., Findling, K. L., Yoshida, T., and Fee, J. A. (1982) Mossbauer Study of a Bacterial Cytochrome-Oxidase - Cytochrome C1aa3 from Thermus-Thermophilus, *Journal of Biological Chemistry* 257, 2489-2492.

- <sup>15</sup> Kiley, P. J., and Beinert, H. (2003) The role of Fe-S proteins in sensing and regulation in bacteria, *Curr Opin Microbiol* 6, 181-185.
- <sup>16</sup> Johnson, M. K. (1998) Iron-sulfur proteins: new roles for old clusters, *Curr Opin Chem Biol* 2, 173-181.
- <sup>17</sup> Drennan, C. L., and Peters, J. W. (2003) Surprising cofactors in metalloenzymes, *Current Opinion in Structural Biology* 13, 220-226.
- <sup>18</sup> Flint, D. H., and Allen, R. M. (1996) Iron-sulfur proteins with nonredox functions, *Chemical Reviews* 96, 2315-2334.
- <sup>19</sup> Fontecave, M., Ollagnier-de-Choudens, S., and Mulliez, E. (2003) Biological radical sulfur insertion reactions, *Chemical Reviews* 103, 2149-2166.
- <sup>20</sup> Zheng, L., White, R. H., Cash, V. L., Jack, R. F., and Dean, D. R. (1993) Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis, *Proc Natl Acad Sci U S A* 90, 2754-2758.
- <sup>21</sup> Zheng, L., Cash, V. L., Flint, D. H., and Dean, D. R. (1998) Assembly of iron-sulfur clusters. Identification of an iscSUA-hscBA-fdx gene cluster from *Azotobacter vinelandii*, *J Biol Chem* 273, 13264-13272.
- <sup>22</sup> Barras, F., Loiseau, L., and Py, B. (2005) How *Escherichia coli* and *Saccharomyces cerevisiae* build Fe/S proteins, *Adv Microb Physiol* 50, 41-101.
- <sup>23</sup> Lill, R., and Muhlenhoff, U. (2005) Iron-sulfur-protein biogenesis in eukaryotes, *Trends in Biochemical Sciences* 30, 133-141.
- <sup>24</sup> Moller, S. G., Kunkel, T., and Chua, N. H. (2001) A plastidic ABC protein involved in intercompartmental communication of light signaling, *Genes & Development* 15, 90-103.
- <sup>25</sup> Xu, X. M., and Moller, S. G. (2004) AtNAP7 is a plastidic SufC-like ATP-binding cassette/ATPase essential for *Arabidopsis* embryogenesis, *Proc Natl Acad Sci U S A* 101, 9143-9148.
- <sup>26</sup> Xu, X. M., Adams, S., Chua, N. H., and Moller, S. G. (2005) AtNAP1 represents an atypical SufB protein in *Arabidopsis* plastids, *Journal of Biological Chemistry* 280, 6648-6654.
- <sup>27</sup> Hjorth, E., Hadfi, K., Zauner, S., and Maier, U. G. (2005) Unique genetic compartmentalization of the SUF system in cryptophytes and characterization of a SufD mutant in *Arabidopsis thaliana*, *Febs Letters* 579, 1129-1135.
- <sup>28</sup> Nachin, L., El Hassouni, M., Loiseau, L., Expert, D., and Barras, F. (2001) SoxR-dependent response to oxidative stress and virulence of *Erwinia chrysanthemi*: the key role of SufC, an orphan ABC ATPase, *Mol Microbiol* 39, 960-972.

- <sup>29</sup> Nachin, L., Loiseau, L., Expert, D., and Barras, F. (2003) SufC: an unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress, *Embo J* 22, 427-437.
- <sup>30</sup> Zheng, M., Wang, X., Templeton, L. J., Smulski, D. R., LaRossa, R. A., and Storz, G. (2001) DNA microarray-mediated transcriptional profiling of the Escherichia coli response to hydrogen peroxide, *J Bacteriol* 183, 4562-4570.
- <sup>31</sup> Outten, F. W., Djaman, O., and Storz, G. (2004) A suf operon requirement for Fe-S cluster assembly during iron starvation in Escherichia coli, *Mol Microbiol* 52, 861-872.
- <sup>32</sup> Loiseau, L., Ollagnier-de Choudens, S., Lascoux, D., Forest, E., Fontecave, M., and Barras, F. (2005) Analysis of the heteromeric CsdA-CsdE cysteine desulfurase, assisting Fe-S cluster biogenesis in Escherichia coli, *J Biol Chem* 280, 26760-26769.
- <sup>33</sup> Ollagnier-de Choudens, S., Loiseau, L., Sanakis, Y., Barras, F., and Fontecave, M. (2005) Quinolinate synthetase, an iron-sulfur enzyme in NAD biosynthesis, *FEBS Lett* 579, 3737-3743.
- <sup>34</sup> Patzer, S. I., and Hantke, K. (1999) SufS is a NifS-like protein, and SufD is necessary for stability of the [2Fe-2S] FhuF protein in Escherichia coli, *J Bacteriol* 181, 3307-3309.
- <sup>35</sup> Takahashi, Y., and Tokumoto, U. (2002) A third bacterial system for the assembly of iron-sulfur clusters with homologs in archaea and plastids, *J Biol Chem* 277, 28380-28383.
- <sup>36</sup> Hantke, K. (2002) Members of the Fur protein family regulate iron and zinc transport in E. coli and characteristics of the Fur-regulated fhuF protein, *Journal of molecular microbiology and biotechnology* 4, 217-222.
- <sup>37</sup> Lee, J. H., Yeo, W. S., and Roe, J. H. (2004) Induction of the sufA operon encoding Fe-S assembly proteins by superoxide generators and hydrogen peroxide: involvement of OxyR, IHF and an unidentified oxidant-responsive factor, *Molecular microbiology* 51, 1745-1755.
- <sup>38</sup> McHugh, J. P., Rodriguez-Quinones, F., Abdul-Tehrani, H., Svistunenko, D. A., Poole, R. K., Cooper, C. E., and Andrews, S. C. (2003) Global iron-dependent gene regulation in Escherichia coli - A new mechanism for iron homeostasis, *Journal of Biological Chemistry* 278, 29478-29486.
- <sup>39</sup> Ollagnier-de Choudens, S., Nachin, L., Sanakis, Y., Loiseau, L., Barras, F., and Fontecave, M. (2003) SufA from Erwinia chrysanthemi. Characterization of a scaffold protein required for iron-sulfur cluster assembly, *J Biol Chem* 278, 17993-18001.
- <sup>40</sup> Mihara, H., Kurihara, T., Yoshimura, T., Soda, K., and Esaki, N. (1997) Cysteine sulfinase, a NIFS-like protein of Escherichia coli with selenocysteine lyase and cysteine desulfurase activities. Gene cloning, purification, and characterization of a novel pyridoxal enzyme, *J Biol Chem* 272, 22417-22424.
- <sup>41</sup> Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M., and Barras, F. (2003) Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase, *J Biol Chem* 278, 38352-38359.

- <sup>42</sup> Outten, F. W., Wood, M. J., Munoz, F. M., and Storz, G. (2003) The SufE protein and the SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in *Escherichia coli*, *J Biol Chem* 278, 45713-45719.
- <sup>43</sup> <sup>43</sup> Mihara, H., Kurihara, T., Yoshimura, T., and Esaki, N. (2000) Kinetic and mutational studies of three NifS homologs from *Escherichia coli*: mechanistic difference between L-cysteine desulfurase and L-selenocysteine lyase reactions, *J Biochem (Tokyo)* 127, 559-567.
- <sup>44</sup> Ollagnier-De-Choudens, S., Lascoux, D., Loiseau, L., Barras, F., Forest, E., and Fontecave, M. (2003) Mechanistic studies of the SufS-SufE cysteine desulfurase: evidence for sulfur transfer from SufS to SufE, *Febs Letters* 555, 263-267.
- <sup>45</sup> Fujii, T., Maeda, M., Mihara, H., Kurihara, T., Esaki, N., and Hata, Y. (2000) Structure of a NifS homologue: X-ray structure analysis of CsdB, an *Escherichia coli* counterpart of mammalian selenocysteine lyase, *Biochemistry* 39, 1263-1273.
- <sup>46</sup> Lima, C. D. (2002) Analysis of the *E. coli* NifS CsdB protein at 2.0 Å reveals the structural basis for perselenide and persulfide intermediate formation, *J Mol Biol* 315, 1199-1208.
- <sup>47</sup> Mihara, H., and Esaki, N. (2002) Bacterial cysteine desulfurases: their function and mechanisms, *Appl Microbiol Biotechnol* 60, 12-23.
- <sup>48</sup> Tirupati, B., Vey, J. L., Drennan, C. L., and Bollinger, J. M. (2004) Kinetic and structural characterization of Slr0077/SufS, the essential cysteine desulfurase from *Synechocystis* sp PCC 6803, *Biochemistry* 43, 12210-12219.
- <sup>49</sup> Ollagnier-de-Choudens, S., Mattioli, T., Takahashi, Y., and Fontecave, M. (2001) Iron-sulfur cluster assembly: characterization of IscA and evidence for a specific and functional complex with ferredoxin, *J Biol Chem* 276, 22604-22607.
- <sup>50</sup> Wu, G., Mansy, S. S., Hemann, C., Hille, R., Surerus, K. K., and Cowan, J. A. (2002) Iron-sulfur cluster biosynthesis: characterization of *Schizosaccharomyces pombe* Isa1, *J Biol Inorg Chem* 7, 526-532.
- <sup>51</sup> Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M., and Barras, F. (2003) Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase, *J Biol Chem* 278, 38352-38359.
- <sup>52</sup> Ollagnier-de-Choudens, S., Sanakis, Y., and Fontecave, M. (2004) SufA/IscA: reactivity studies of a class of scaffold proteins involved in Fe-S cluster assembly, *Journal of Biological Inorganic Chemistry* 9, 828-838.
- <sup>53</sup> Krebs, C., Agar, J. N., Smith, A. D., Frazzon, J., Dean, D. R., Huynh, B. H., and Johnson, M. K. (2001) IscA, an alternate scaffold for Fe-S cluster biosynthesis, *Biochemistry* 40, 14069-14080.
- <sup>54</sup> Wollenberg, M., Berndt, C., Bill, E., Schwenn, J. D., and Seidler, A. (2003) A dimer of the FeS cluster biosynthesis protein IscA from cyanobacteria binds a [2Fe2S] cluster between two



protomers and transfers it to [2Fe2S] and [4Fe4S] apo proteins, *Eur J Biochem* 270, 1662-1671.

<sup>55</sup> Agar, J. N., Krebs, C., Frazzon, J., Huynh, B. H., Dean, D. R., and Johnson, M. K. (2000) IscU as a scaffold for iron-sulfur cluster biosynthesis: sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU, *Biochemistry* 39, 7856-7862.

<sup>56</sup> Smith, A. D., Agar, J. N., Johnson, K. A., Frazzon, J., Amster, I. J., Dean, D. R., and Johnson, M. K. (2001) Sulfur transfer from IscS to IscU: the first step in iron-sulfur cluster biosynthesis, *J Am Chem Soc* 123, 11103-11104.

<sup>57</sup> Jensen, L. T., and Culotta, V. C. (2000) Role of *Saccharomyces cerevisiae* ISA1 and ISA2 in iron homeostasis, *Mol Cell Biol* 20, 3918-3927.

<sup>58</sup> Wu, S. P., and Cowan, J. A. (2003) Iron-sulfur cluster biosynthesis. A comparative kinetic analysis of native and Cys-substituted ISA-mediated [2Fe-2S]<sub>2</sub><sup>+</sup> cluster transfer to an apoferritin target, *Biochemistry* 42, 5784-5791.

<sup>59</sup> Popescu, C. V., Bates, D. M., Beinert, H., Munck, E., and Kiley, P. J. (1998) Mossbauer spectroscopy as a tool for the study of activation/inactivation of the transcription regulator FNR in whole cells of *Escherichia coli*, *Proceedings of the National Academy of Sciences of the United States of America* 95, 13431-13435.

<sup>60</sup> Benda, R., Bui, B. T. S., Schunemann, V., Florentin, D., Marquet, A., and Trautwein, A. X. (2002) Iron-sulfur clusters of biotin synthase in vivo: A Mossbauer Study, *Biochemistry* 41, 15000-15006.

<sup>61</sup> Ding, H., and Clark, R. J. (2004) Characterization of iron binding in IscA, an ancient iron-sulphur cluster assembly protein, *Biochem J* 379, 433-440.

<sup>62</sup> Ding, H., Clark, R. J., and Ding, B. (2004) IscA mediates iron delivery for assembly of iron-sulfur clusters in IscU under the limited accessible free iron conditions, *J Biol Chem* 279, 37499-37504.

<sup>63</sup> Morimoto, K., Yamashita, E., Kondou, Y., Lee, S. J., Arisaka, F., Tsukihara, T., and Nakai, M. (2006) The Asymmetric IscA Homodimer with an Exposed [2Fe-2S] Cluster Suggests the Structural Basis of the Fe-S Cluster Biosynthetic Scaffold, *J Mol Biol*.

<sup>64</sup> Fontecave, M., Choudens, S. O., Py, B., and Barras, F. (2005) Mechanisms of iron-sulfur cluster assembly: the SUF machinery, *J Biol Inorg Chem*, 1-9.

<sup>65</sup> Urbina, H. D., Silberg, J. J., Hoff, K. G., and Vickery, L. E. (2001) Transfer of sulfur from IscS to IscU during Fe/S cluster assembly, *J Biol Chem* 276, 44521-44526.

<sup>66</sup> Venter, J. C. (2001). "The sequence of the human genome (vol 292, pg 1304, 2001)." *Science* 292(5523): 1838-1838.

- <sup>67</sup> Walters, K. J., Ferentz, A. E., Hare, B. J., Hidalgo, P., Jasanoff, A., Matsuo, H., and Wagner, G. (2001) Characterizing protein-protein complexes and oligomers by nuclear magnetic resonance spectroscopy, *Methods in enzymology* **339**, 238-258.
- <sup>68</sup> Kelly, G., Prasannan, S., Daniell, S., Fleming, K., Frankel, G., Dougan, G., Connerton, I., and Matthews, S. (1999) Structure of the cell-adhesion fragment of intimin from enteropathogenic *Escherichia coli*, *Nature structural biology* **6**, 313-318.
- <sup>69</sup> Kelly, G., S. Prasannan, et al. (1999). "Structure of the cell-adhesion fragment of intimin from enteropathogenic *Escherichia coli*." *Nature Structural Biology* **6**(4): 313-318.
- <sup>70</sup> Garrett, D. S., Y. J. Seok, et al. (1997). "Solution structure of the 30 kDa N-terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR." *Biochemistry* **36**(9): 2517-2530.
- <sup>71</sup> Kelly, M. J. S., L. J. Ball, et al. (2001). "The NMR structure of the 47-kDa dimeric enzyme 3,4-dihydroxy-2-butanone-4-phosphate synthase and ligand binding studies reveal the location of the active site." *Proceedings of the National Academy of Sciences of the United States of America* **98**(23): 13025-13030
- <sup>72</sup> Abragam, A. (1961) Principles of Nuclear Magnetism. (Oxford) Clarendon Press
- <sup>73</sup> Williamson, M. P., T. F. Havel, et al. (1985). "Solution Conformation of Proteinase Inhibitor-Iia from Bull Seminal Plasma by H-1 Nuclear Magnetic-Resonance and Distance Geometry." *Journal of Molecular Biology* **182**(2): 295-315.
- <sup>74</sup> Xu, R., B. Ayers, et al. (1999). "Chemical ligation of folded recombinant proteins: Segmental isotopic labeling of domains for NMR studies." *Proceedings of the National Academy of Sciences of the United States of America* **96**(2): 388-393.
- <sup>75</sup> Browne, D. T., G. L. Kenyon, et al. (1973). "Studies of Macromolecular Structure by C-13 Nuclear Magnetic-Resonance .2. A Specific Labeling Approach to Study of Histidine Residues in Proteins." *Journal of the American Chemical Society* **95**(4): 1316-1323.
- <sup>76</sup> Venters, R. A., B. T. Farmer, et al. (1996). "Characterizing the use of perdeuteration in NMR studies of large proteins C-13, N-15 and H-1 assignments of human carbonic anhydrase II." *Journal of Molecular Biology* **264**(5): 1101-1116.
- <sup>77</sup> Gardner, K. H. and L. E. Kay (1998). "The use of H-2, C-13, N-15 multidimensional NMR to study the structure and dynamics of proteins." *Annual Review of Biophysics and Biomolecular Structure* **27**: 357-406.
- <sup>78</sup> Sattler, M. and S. W. Fesik (1996). "Use of deuterium labeling in NMR: Overcoming a sizeable problem." *Structure* **4**(11): 1245-1249.
- <sup>79</sup> Nietlispach, D., R. T. Clowes, et al. (1996). "An approach to the structure determination of larger proteins using triple resonance NMR experiments in conjunction with random fractional deuteration." *Journal of the American Chemical Society* **118**(2): 407-415.

- <sup>80</sup> Hajduk, P. J., D. J. Augeri, et al. (2000). "NMR-based screening of proteins containing C-13-labeled methyl groups." Journal of the American Chemical Society **122**(33): 7898-7904.
- <sup>81</sup> Gardner, K. H., M. K. Rosen, et al. (1997). "Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR." Biochemistry **36**(6): 1389-1401.
- <sup>82</sup> Pervushin, K., R. Riek, et al. (1997). "Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution." Proceedings of the National Academy of Sciences of the United States of America **94**(23): 12366-12371.
- <sup>83</sup> Boisbouvier, J., Blackledge, M., Sollier, A., and Marion, D. (2000) Simultaneous determination of disulphide bridge topology and three-dimensional structure using ambiguous intersulphur distance restraints: possibilities and limitations, *Journal of biomolecular NMR* **16**, 197-208.
- <sup>84</sup> Uhlen, M. and T. Moks (1990). "Gene Fusions for Purpose of Expression - an Introduction." Methods in Enzymology **185**: 129-143.
- <sup>85</sup> Kohno, T., H. Kusunoki, et al. (1998). "A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin fusion system: An application to the production of mastoparan-X uniformly enriched with N-15 and N-15/C-13." Journal of Biomolecular Nmr **12**(1): 109-121.
- <sup>86</sup> Mannaitis, T., Frisch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual (New York) Cold Spring Harbour Laboratory.
- <sup>87</sup> Jansson, M., Li, Y. C., Jendeberg, L., Anderson, S., Montelione, B. T., and Nilsson, B. (1996) High-level production of uniformly 15N- and 13C-enriched fusion proteins in Escherichia coli, *Journal of biomolecular NMR* **7**, 131-141.
- <sup>88</sup> Jansson, M., Y. C. Li, et al. (1996). "High-level production of uniformly N-15- and C-13-enriched fusion proteins in Escherichia coli." Journal of Biomolecular Nmr **7**(2): 131-141.
- <sup>89</sup> Reilly, D., and Fairbrother, W. J. (1994) A novel isotope labeling protocol for bacterially expressed proteins, *Journal of biomolecular NMR* **4**, 459-462.
- <sup>90</sup> Katz, J. J. and H. L. Crespi (1966). "Deuterated Organisms - Cultivation and Uses." Science **151**(3715): 1187-&.
- <sup>91</sup> Wüthrich, K. (1986) NMR of Protein and Nucleic Acids. (New York) Wiley & Sons.
- <sup>92</sup> Sattler, M., J. Schleucher, et al. (1999). "Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients." Progress in Nuclear Magnetic Resonance Spectroscopy **34**(2): 93-158.
- <sup>93</sup> Yang, D. W. and L. E. Kay (1999). "TROSY triple-resonance four-dimensional NMR spectroscopy of a 46 ns tumbling protein." Journal of the American Chemical Society **121**(11): 2571-2575.

- <sup>94</sup> Szyperski, T., Wider, G., Bushweller, J. H., and Wuthrich, K. (1993) 3D <sup>13</sup>C-<sup>15</sup>N-heteronuclear two-spin coherence spectroscopy for polypeptide backbone assignments in <sup>13</sup>C-<sup>15</sup>N-double-labeled proteins, *Journal of biomolecular NMR* **3**, 127-132.
- <sup>95</sup> Szyperski, T., G. Wider, et al. (1993). "Reduced Dimensionality in Triple-Resonance Nmr Experiments." *Journal of the American Chemical Society* **115**(20): 9307-9308.
- <sup>96</sup> Brutscher, B., J. P. Simorre, et al. (1994). "Design of a Complete Set of 2-Dimensional Triple-Resonance Experiments for Assigning Labeled Proteins." *Journal of Magnetic Resonance Series B* **105**(1): 77-82.
- <sup>97</sup> Zimmerman, D. E., C. A. Kulikowski, et al. (1997). "Automated analysis of protein NMR assignments using methods from artificial intelligence." *Journal of Molecular Biology* **269**(4): 592-610.
- <sup>98</sup> Morelle, N., B. Brutscher, et al. (1995). "Computer Assignment of the Backbone Resonances of Labeled Proteins Using 2-Dimensional Correlation Experiments." *Journal of Biomolecular Nmr* **5**(2): 154-160.
- <sup>99</sup> Johnson, B. A., and Blevins, R.A. (1994), *Journal of Biomolecular NMR*, **4**:603-614.
- <sup>100</sup> Cavanagh, J., Fairbrother, W.J., Skelton, H.J., et Palmer, A.G. 3<sup>rd</sup> (1996). "ProteinNMR Spectroscopy – Principles and Practices. In anonymous Academic Press Inc.
- <sup>101</sup> Grzesiek, S., J. Anglister, et al. (1993). "Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in C-13/N-15-Enriched Proteins by Isotropic Mixing of C-13 Magnetization." *Journal of Magnetic Resonance Series B* **101**(1): 114-119.
- <sup>102</sup> Logan, T. M., E. T. Olejniczak, et al. (1993). "A General-Method for Assigning Nmr-Spectra of Denatured Proteins Using 3d Hc(Co)Nh-Tocsy Triple Resonance Experiments." *Journal of Biomolecular Nmr* **3**(2): 225-231.
- <sup>103</sup> Montelione, G. T., B. A. Lyons, et al. (1992). "An Efficient Triple Resonance Experiment Using C-13 Isotropic Mixing for Determining Sequence-Specific Resonance Assignments of Isotopically-Enriched Proteins." *Journal of the American Chemical Society* **114**(27): 10974-10975.
- <sup>104</sup> Kay, L. E., G. Y. Xu, et al. (1993). "A Gradient-Enhanced Hcch Tocsy Experiment for Recording Side-Chain H-1 and C-13 Correlations in H2o Samples of Proteins." *Journal of Magnetic Resonance Series B* **101**(3): 333-337.
- <sup>105</sup> Majumdar, A., H. Wang, et al. (1993). "Sensitivity Improvement in 2d and 3d Hcch Spectroscopy Using Heteronuclear Cross-Polarization." *Journal of Biomolecular Nmr* **3**(4): 387-397.
- <sup>106</sup> Farmer, B. T. and R. A. Venters (1996). "Assignment of aliphatic side-chain H-1(N)/N-15 resonances in perdeuterated proteins." *Journal of Biomolecular Nmr* **7**(1): 59-71.

- <sup>107</sup> Jeener, J., B. H. Meier, et al. (1979). "Investigation of Exchange Processes by 2-Dimensional Nmr-Spectroscopy." Journal of Chemical Physics **71**(11): 4546-4553.
- <sup>108</sup> Greenfield, N. J., Huang, Y. J., Palm, T., Swapna, G. V., Monleon, D., Montelione, G. T., and Hitchcock-DeGregori, S. E. (2001) Solution NMR structure and folding dynamics of the N terminus of a rat non-muscle alpha-tropomyosin in an engineered chimeric protein, *Journal of molecular biology* **312**, 833-847.
- <sup>109</sup> Linge, J. P., O'Donoghue, S. I., and Nilges, M. (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA, *Methods in enzymology* **339**, 71-90.
- <sup>110</sup> Karplus, M. (1959). "Contact Electron-Spin Coupling of Nuclear Magnetic Moments." Journal of Chemical Physics **30**(1): 11-15.
- <sup>111</sup> Cornilescu, G., J. S. Hu, et al. (1999). "Identification of the hydrogen bonding network in a protein by scalar couplings." Journal of the American Chemical Society **121**(12): 2949-2950.
- <sup>112</sup> Cornilescu, G., B. E. Ramirez, et al. (1999). "Correlation between (3h)J(NC ' ) and hydrogen bond length in proteins." Journal of the American Chemical Society **121**(26): 6275-6279.
- <sup>113</sup> Wishart, D. S. and B. D. Sykes (1994). Chemical-Shifts as a Tool for Structure Determination. Nuclear Magnetic Resonance, Pt C. **239**: 363-392.
- <sup>114</sup> Wishart, D. S., B. D. Sykes, et al. (1991). "Relationship between Nuclear-Magnetic-Resonance Chemical-Shift and Protein Secondary Structure." Journal of Molecular Biology **222**(2): 311-333.
- <sup>115</sup> Wishart, D. S. and B. D. Sykes (1994). "The C-13 Chemical-Shift Index - a Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data." Journal of Biomolecular Nmr **4**(2): 171-180.
- <sup>116</sup> Cornilescu, G., F. Delaglio, et al. (1999). "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." Journal of Biomolecular Nmr **13**(3): 289-302.
- <sup>117</sup> Clore, G. M., Gronenborn, A. M., and Bax, A. (1998) A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information, *J Magn Reson* **133**, 216-221.
- <sup>118</sup> Losonczi, J. A., Andrec, M., Fischer, M. W., and Prestegard, J. H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition, *J Magn Reson* **138**, 334-342.
- <sup>119</sup> Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data, *Journal of biomolecular NMR* **16**, 23-28.

- <sup>120</sup> Tjandra, N. and A. Bax (1997). "Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium (vol 278, pg 1111, 1997)." Science **278**(5344): 1697-1697.
- <sup>121</sup> Hus, J. C., D. Marion, et al. (2001). "Determination of protein backbone structure using only residual dipolar couplings." Journal of the American Chemical Society **123**(7): 1541-1542.
- <sup>122</sup> Tolman, J. R., J. M. Flanagan, et al. (1995). "Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information for Structure Determination in Solution." Proceedings of the National Academy of Sciences of the United States of America **92**(20): 9279-9283.
- <sup>123</sup> Gochin, M. and H. Roder (1995). "Protein-Structure Refinement Based on Paramagnetic Nmr Shifts - Applications to Wild-Type and Mutant Forms of Cytochrome-C." Protein Science **4**(2): 296-305.
- <sup>124</sup> Boisbouvier, J., P. Gans, et al. (1999). "Long-range structural information in NMR studies of paramagnetic molecules from electron spin-nuclear spin cross-correlated relaxation." Journal of the American Chemical Society **121**(33): 7700-7701.
- <sup>125</sup> Clore, G. M., Starich, M. R., and Gronenborn, A. M. (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses, *Journal of the American Chemical Society* **120**, 10571-10572.
- <sup>126</sup> Hansen, M. R., Mueller, L., and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions, *Nature structural biology* **5**, 1065-1074.
- <sup>127</sup> Zweckstetter, M., and Bax, A. (2001) Characterization of molecular alignment in aqueous suspensions of Pfl bacteriophage, *Journal of biomolecular NMR* **20**, 365-377.
- <sup>128</sup> Nilges, M. (1995) Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities, *Journal of molecular biology* **245**, 645-660.
- <sup>129</sup> Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. (2003) ARIA: automated NOE assignment and NMR structure calculation, *Bioinformatics (Oxford, England)* **19**, 315-316.
- <sup>130</sup> Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta crystallographica* **54**, 905-921.
- <sup>131</sup> Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M., and Bax, A. (1997) Use of dipolar <sup>1</sup>H-<sup>15</sup>N and <sup>1</sup>H-<sup>13</sup>C couplings in the structure determination of magnetically oriented macromolecules in solution, *Nature structural biology* **4**, 732-738.

- <sup>132</sup> Cupp-Vickery, J. R., Silberg, J. J., Ta, D. T., and Vickery, L. E. (2004) Crystal structure of IscA, an iron-sulfur cluster assembly protein from *Escherichia coli*, *Journal of molecular biology* 338, 127-137.
- <sup>133</sup> Mittard, V., Morelle, N., Brutscher, B., Simorre, J. P., Marion, D., Stein, M., Jacquot, J. P., Lirsac, P. N., and Lancelin, J. M. (1995) <sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N-NMR resonance assignments of oxidized thioredoxin h from the eukaryotic green alga *Chlamydomonas reinhardtii* using new methods based on two-dimensional triple-resonance NMR spectroscopy and computer-assisted backbone assignment, *European journal of biochemistry / FEBS* 229, 473-485.
- <sup>134</sup> Wishart, D. S., and Sykes, B. D. (1994) The <sup>13</sup>C chemical-shift index: a simple method for the identification of protein secondary structure using <sup>13</sup>C chemical-shift data, *Journal of biomolecular NMR* 4, 171-180.
- <sup>135</sup> Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Progress in Nuclear Magnetic Resonance Spectroscopy* 34, 93-158.
- <sup>136</sup> Van Melckebeke, H., Simorre, J. P., and Brutscher, B. (2004) Amino acid-type edited NMR experiments for methyl-methyl distance measurement in <sup>13</sup>C-labeled proteins, *Journal of the American Chemical Society* 126, 9584-9591.
- <sup>137</sup> Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., and Wuthrich, K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids--IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy, *European journal of biochemistry / FEBS* 256, 1-15.
- <sup>138</sup> Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *Journal of biomolecular NMR* 6, 277-293.
- <sup>139</sup> Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *Journal of biomolecular NMR* 13, 289-302.
- <sup>140</sup> Bodenhausen, G., and Ernst, R. R. (1981) The Accordion Experiment, a Simple Approach to 3-Dimensional Nmr-Spectroscopy, *Journal of Magnetic Resonance* 45, 367-373.
- <sup>141</sup> Meissner, A., Schulte-Herbruggen, T., and Sorensen, O. W. (1998) Relaxation artifacts and their suppression in multidimensional E.COSY-type NMR experiments for measurement of J coupling constants in C-13- or N-15-labeled proteins, *Journal of the American Chemical Society* 120, 7989-7990.
- <sup>142</sup> Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra, *Journal of Magnetic Resonance* 131, 373-378.

- <sup>143</sup> Brutscher, B. (2001) Accurate measurement of small spin-spin couplings in partially aligned molecules using a novel J-mismatch compensated spin-state-selection filter, *J Magn Reson* **151**, 332-338.
- <sup>144</sup> Dosset, P., Hus, J. C., Marion, D., and Blackledge, M. (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings, *Journal of biomolecular NMR* **20**, 223-231.
- <sup>145</sup> Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993) Main-chain bond lengths and bond angles in protein structures, *Journal of molecular biology* **231**, 1049-1067.
- <sup>146</sup> Wada, K., Hasegawa, Y., Gong, Z., Minami, Y., Fukuyama, K., and Takahashi, Y. (2005) Crystal structure of Escherichia coli SufA involved in biosynthesis of iron-sulfur clusters: implications for a functional dimer, *FEBS letters* **579**, 6543-6548.
- <sup>147</sup> Herrmann, T., Guntert, P., and Wuthrich, K. (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS, *Journal of biomolecular NMR* **24**, 171-189.
- <sup>148</sup> Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information, *Journal of the American Chemical Society* **125**, 1731-1737.
- <sup>149</sup> Morimoto, K., Yamashita, E., Kondou, Y., Lee, S. J., Arisaka, F., Tsukihara, T., and Nakai, M. (2006) The asymmetric IscA homodimer with an exposed [2Fe-2S] cluster suggests the structural basis of the Fe-S cluster biosynthetic scaffold, *Journal of molecular biology* **360**, 117-132.
- <sup>150</sup> Bilder, P. W., Ding, H., and Newcomer, M. E. (2004) Crystal structure of the ancient, Fe-S scaffold IscA reveals a novel protein fold, *Biochemistry* **43**, 133-139.
- <sup>151</sup> Imlay, J. A. (2006) Iron-sulphur clusters and the problem with oxygen, *Molecular microbiology* **59**, 1073-1082.
- <sup>152</sup> M. Lombard, D. Touati, M. Fontecave and V. Nivière, Superoxide reductase as a unique defense system against superoxide stress in the microaerophile *Treponema pallidum*, *J. Biol. Chem.* **275** (2000), pp. 27021–27026.
- <sup>153</sup> F.E. Jenney Jr., M.F.J.M. Verhagen, X. Cui and M.W.W. Adams, Anaerobic microbes: oxygen detoxification without superoxide dismutase, *Science* **286** (1999), pp. 306–309.
- <sup>154</sup> V. Nivière and M. Fontecave, Discovery of superoxide reductase: an historical perspective, *J. Biol. Inorg. Chem.* **9** (2004), pp. 119–123.
- <sup>155</sup> T. Jovanovic, C. Ascenso, K.R.O. Hazlett, R. Sikkink, C. Krebs, R. Litwiller, L.M. Benson, I. Moura, J.J.G. Moura, J.D. Radolf, B.H. Huynh, S. Naylor and F. Rusnak, Neelaredoxin, an iron-binding protein from the syphilis spirochete, *Treponema pallidum*, is a superoxide reductase, *J. Biol. Chem.* **275** (2000), pp. 28439–28448.



- <sup>156</sup> A. Coelho, P. Matias, V. Fülöp, A. Thompson, A. Gonzalez and M.A. Carrondo, Desulfoferrodoxin structure determined by MAD phasing and refinement to 1.9-Å resolution reveals a unique combination of a tetrahedral FeS<sub>4</sub> centre with a square pyramidal FeSN<sub>4</sub> centre, *J. Biol. Inorg. Chem.* **2** (1997), pp. 680–689.
- <sup>157</sup>
- <sup>158</sup> Yeh, A. P., Hu, Y., Jenney, F. E., Jr., Adams, M. W., and Rees, D. C. (2000) Structures of the superoxide reductase from *Pyrococcus furiosus* in the oxidized and reduced states, *Biochemistry* **39**, 2499-2508.
- <sup>159</sup> Santos-Silva, T., Trincao, J., Carvalho, A. L., Bonifacio, C., Auchere, F., Raleiras, P., Moura, I., Moura, J. J., and Romao, M. J. (2006) The first crystal structure of class III superoxide reductase from *Treponema pallidum*, *J Biol Inorg Chem* **11**, 548-558.
- <sup>160</sup> E.D. Coulter, J.P. Emerson, D.M. Kurtz Jr. and D.E. Cabelli, Superoxide reactivity of rubredoxin oxidoreductase (Desulfoferrodoxin) from *Desulfovibrio vulgaris*: a pulse radiolysis study, *J. Am. Chem. Soc.* **122** (2000), pp. 11555–11556.
- <sup>161</sup> M. Lombard, C. Houée-Levin, D. Touati, M. Fontecave and V. Nivière, Superoxide reductase from *Desulfoarculus baarsii*: reaction mechanism and role of glutamate 47 and lysine 48 in catalysis, *Biochemistry* **40** (2001), pp. 5032–5040.
- <sup>162</sup> V. Nivière, M. Lombard, M. Fontecave and C. Houée-Levin, Pulse radiolysis studies on superoxide reductase from *Treponema pallidum*, *FEBS Lett.* **497** (2001), pp. 171–173.
- <sup>163</sup> I.A. Abreu, L.M. Saraiva, C.M. Soares, M. Teixeira and D.E. Cabelli, The mechanism of superoxide scavenging by *Archaeoglobus fulgidus* neelaredoxin, *J. Biol. Chem.* **276** (2001), pp. 38995–39001.
- <sup>164</sup> J.P. Emerson, E.D. Coulter, D.E. Cabelli, R.S. Phillips and D.M. Kurtz Jr., Kinetics and mechanism of superoxide reduction by two-iron superoxide reductase from *Desulfovibrio vulgaris*, *Biochemistry* **41** (2002), pp. 4348–4357.
- <sup>165</sup> V. Nivière, M. Asso, C.O. Weill, M. Lombard, B. Guigliarelli, V. Favaudon and C. Houée-Levin, Superoxide reductase from *Desulfoarculus baarsii*: identification of protonation steps in the enzymatic mechanism, *Biochemistry* **43** (2004), pp. 808–818.
- <sup>166</sup> Huber, J. G., Gaillard, J. & Moulis, J. M. Nmr of Chromatium-Vinosum Ferredoxin - Evidence for Structural Inequivalence and Impeded Electron-Transfer between the 2 4fe-4s Clusters. *Biochemistry* **34**, 194-205 (1995).
- <sup>167</sup> Bax, A. & Davis, D. G. Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. *Journal of Magnetic Resonance* **65**, 355-360 (1985).
- <sup>168</sup> Macura, S., Huang, Y., Suter, D. & Ernst, R. R. Two-Dimensional Chemical-Exchange and Cross-Relaxation Spectroscopy of Coupled Nuclear Spins. *Journal of Magnetic Resonance* **43**, 259-281 (1981).

- <sup>169</sup> States, D. J., Haberkorn, R. A. & Ruben, D. J. A Two-Dimensional Nuclear Overhauser Experiment with Pure Absorption Phase in 4 Quadrants. *Journal of Magnetic Resonance* **48**, 286-292 (1982).
- <sup>170</sup> Lombard, M., Touati, D., Fontecave, M., and Niviere, V. (2000) Superoxide reductase as a unique defense system against superoxide stress in the microaerophile *Treponema pallidum*, *Journal of Biological Chemistry* **275**, 27021-27026.
- <sup>171</sup> Lombard, M., Houee-Levin, C., Touati, D., Fontecave, M., and Niviere, V. (2001) Superoxide reductase from *Desulfoarculus baarsii*: reaction mechanism and role of glutamate 47 and lysine 48 in catalysis, *Biochemistry* **40**, 5032-5040.
- <sup>172</sup> Brutsher, B., Cordier, F., Simorre, J. P., Caffrey, M., and Marion, D., High-resolution 3D HNCOCOA experiment applied to a 28 Kda paramagnetic protein. *Journal of biomolecular NMR* **5**, 202-6 (1995)
- <sup>173</sup> Santos-Silva, T., Trincao, J., Carvalho, A. L., Bonifacio, C., Auchere, F., Moura, I., Moura, J. J. G., and Romao, M. J. (2005) Superoxide reductase from the syphilis spirochete *Treponema pallidum*: crystallization and structure determination using soft X-rays, *Acta Crystallographica Section F-Structural Biology and Crystallization Communications* **61**, 967-970.
- <sup>174</sup> Bradford, M. M. (1976) Rapid and Sensitive Method for Quantitation of Microgram Quantities of Protein Utilizing Principle of Protein-Dye Binding, *Analytical Biochemistry* **72**, 248-254.
- <sup>175</sup> 1. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *Journal of molecular graphics* **14**, 51-55, 29-32.
- <sup>176</sup> Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M., and Nilges, M. (2003) Refinement of protein structures in explicit solvent, *Proteins* **50**, 496-506.
- <sup>177</sup> Engh, R. A., and Huber, R. (1991) Accurate Bond and Angle Parameters for X-Ray Protein-Structure Refinement, *Acta Crystallographica Section A* **47**, 392-400.